

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 14:26:07 ; Search time 1450.75 Seconds  
(Without alignments)  
2945.214 Million cell updates/sec

Title: US-09-587-653-1

Sequence: 1 aguuugcgaccacuuugugga.....ccccuuugaaauaaacuu 259

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%  
Listing first 45 summaries

Database :  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_ov:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_com:\*  
20: em\_or:\*  
21: em\_ov:\*  
22: em\_pat:\*  
23: em\_ph:\*  
24: em\_pl:\*  
25: em\_ro:\*  
26: em\_sts:\*  
27: em\_sy:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_htgo\_hum:\*  
31: em\_htgo\_inv:\*  
32: em\_htgo\_rtd:\*  
33: em\_htg\_hum:\*  
34: em\_htg\_inv:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	259	100.0	9399	6	AX057392	AX057392 Sequence
2	259	100.0	9399	14	AF179612	AF179612 Hepatitis
3	257.4	99.4	259	6	AX055778	AX055778 Sequence
4	257.4	99.4	259	6	AX055779	AX055779 Sequence
5	257.4	99.4	259	6	AX055780	AX055780 Sequence
6	257.4	99.4	259	6	AX055781	AX055781 Sequence
7	257.4	99.4	309	6	AX055777	AX055777 Sequence
8	257.4	99.4	357	14	HGB18973	Y18973 Hepatitis G
9	257.4	99.4	9397	14	HGB277947	AJ277947 Hepatitis
10	82	31.7	82	6	AX055790	AX055790 Sequence
11	33.2	12.8	185516	9	AC004485	AC004485 Homo sapi
12	33.2	12.6	289563	2	AL450993	AL450993 Homo sapi
13	32.6	12.6	172751	10	CNS076CP	AL591826 BAC 13C18
14	32.4	12.5	2354	14	PPM76VPI	D00056 Monkey B-1y
15	32.4	12.5	2354	14	PPM76VPI	M14494 Monkey B-1y
16	32.2	12.4	185273	2	AC019038	AC019038 Homo sapi
17	32	12.4	3179	6	AX179748	AX179748 Sequence
18	31.8	12.3	185302	2	AC010177	AC010177 Homo sapi
19	31.8	12.3	191877	9	AC007207	AC007207 Homo sapi
20	31.8	12.3	210004	9	AC069240	AC069240 Homo sapi
21	31.6	12.2	4498	3	AB030450	AB030450 Drosophila
22	31.4	12.1	1257	4	OCIGHENH	X13700 Rabbit IgH
23	31.4	12.1	137023	2	AC091091	AC091091 Gallus ga
24	31	12.0	177990	2	AC016385	AC016385 Homo sapi
25	31	12.0	205803	2	AC015802	AC015802 Homo sapi
26	31	12.0	206400	10	AC087184	AC087184 Mus muscu
27	30.8	11.9	2354	14	PPM76VPI	L29432 Monkey B-1y
28	30.8	11.9	5089	14	PPM76VPI	D00055 Monkey B-1y
29	30.8	11.9	5089	14	PPM76VPI	M30540 Monkey B-1y
30	30.8	11.9	5270	14	PPL	K02562 Lymphotropl
31	30.8	11.9	84904	9	AP001051	AP001051 Homo sapi
32	30.8	11.9	340000	9	AP001751	AP001751 Homo sapi
33	30.6	11.8	37762	3	AC005836	AC005836 Leishmani
34	30.6	11.8	159030	2	AC091729	AC091729 Homo sapi
35	30.6	11.8	172438	2	AC015711	AC015711 Homo sapi
36	30.6	11.8	180359	2	AC022064	AC022064 Homo sapi
37	30.6	11.8	185255	2	AC068181	AC068181 Homo sapi
38	30.6	11.8	219218	2	AC020817	AC020817 Mus muscu
39	30.4	11.7	21420	1	SCF85	AL10470 Streptomy
40	30.4	11.7	154279	2	AL138781	AL138781 Homo sapi
41	30.4	11.7	189121	2	AC02285	AC02285 Homo sapi
42	30.4	11.7	193335	2	AC016826	AC016826 Homo sapi
43	30.2	11.7	71578	2	AC025333	AC025333 Homo sapi
44	30.2	11.7	158519	9	AL135926	AL135926 Human DNA
45	30.2	11.7	162264	2	AC023787	AC023787 Homo sapi

## ALIGNMENTS

RESULT 1  
LOCUS AX057392 9399 bp DNA  
DEFINITION Sequence 1 from Patent WO00/5337.  
VERSION AX057392.1 GI:12310132  
SOURCE  
ORGANISM Hepatitis GB virus B.  
Hepatitis GB virus B.  
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.  
REFERENCE 1 (Bases 1 to 9399)  
AUTHORS Bukh, J., Yanagisaki, M., Emerson, S.U. and Purcell, R.H.  
TITLE Infectious cDNA clone of gb virus b and uses thereof  
JOURNAL Patient: WO 0075337-A, 1 14-DEC-2000:  
THE GOVERNMENT OF THE UNITED STATES OF AMERICA (US)  
FEATURES  
SOURCE Location/Qualifiers  
1..9399  
/organism="Hepatitis GB virus B"  
/db\_xref="taxon:39113"  
BASE COUNT 2125 a 2326 c 2430 g 2518 t



DEFINITION	Sequence 2 from Patent WO0073466.
ACCESSION	AX055778
VERSION	AX055778.1
KEYWORDS	GI:12228690
SOURCE	unidentified.
ORGANISM	unidentified
REFERENCE	unclassified.
AUTHORS	1 (bases 1 to 259)
TITLE	Traboni,C.I.
JOURNAL	Novel gbv sequence Patent: WO 0073466-A 2 07-DEC-2000; ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE P. ANGELETTI, S.P.A. (It)
FEATURES	
source	Location/Qualifiers 1..259 /organism="unidentified" /db_xref="taxon:32644" /note="GBV-B-like virus."
BASE COUNT	53 a 64 c 81 g 61 t
ORIGIN	

Query Match	99.48%	Score 257.4	DB 6	Length 259
Best Local Similarity	76.18%	Pred. No. 2666		
Matches 197	61	Mismatches	1	Indels 0
				Gaps 0

[illegible]

RESULT	4
AX05779/c	
LOCUS	
AX05779	259 bp
	mRNA
	PAT
	13-JAN-2001

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
AX055779	AX055779.1	GI:12228891		unidentified.	unidentified.			
				unclassified.	unclassified.			
				1 (bases 1 to 259)	1 (bases 1 to 259)			
				Traboni,C.I.	Traboni,C.I.			
				Novel gpbv sequence	Novel gpbv sequence			
				Patent: WO 007346-A 3	Patent: WO 007346-A 3			
				07-DEC-2000;	07-DEC-2000;			
				ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE	ISTITUTO DI RICERCHE DI BIOLOGIA MOLECOLARE			
				P. ANGELETTI; S.P.A	P. ANGELETTI; S.P.A			
				(IT)	(IT)			

FEATURES	SOURCE	location/Qualifiers
		1. 259
		/organism="unidentified"
		/db_xref="taxon:32644"
		/note=GBV-B-like virus"
BASE COUNT		61 a 81 c 64 g 5
ORIGIN		

Query Match	99.48;	Score 257.4;	DB 6;	Length 259;
-------------	--------	--------------	-------	-------------

Best Local Similarity 76.1%; Pred. No. 2e-66;  
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;

[illegible]

RESULT	5		
AX055780			
LOCUS	AX055780	259 bp	DNA
DEFINITION	Sequence	4 from Patent WO0073466.	PAT
ACCESSION	AX055780		13-JAN-2001
VERSION	AX055780.1	GI:12228892	
KEYWORDS	.		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	unclassified.		
AUTHORS	1 (bases 1 to 259)		
TITLE	Traboni,C.I.		
JOURNAL	Novel gbu sequence		
	Patent: WO 0073466-A 4		
	07-DEC-2000;		

```

FEATURES
    source                location/Qualifiers
                        1..259
                        /organism="unidentified"
                        /db_xref="taxon:32644"
                        /note="GBV-B-like virus"
    3' UTR                1..259
                        /note="cDNA of positive strand 3' UTR of GBV-B"
    BASE COUNT            53 a      64 c      81 g      61 t
    ORIGIN

```

[illegible]

[illegible]

BASE COUNT	ORIGIN	Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
77 a	88 c	101 g	91 t	99.4%: 76.1%: Pred. No. 1.9e-66; Mismatches 1; Indels 0; Gaps 0.	257.4;	DB 6;	309;		
3' UTR		1. .309	/note="GBV-B-like virus"						
		/note="cDNA complementary to the last 49 published 3'UTR nucleotides of GBV-B plus a novel nucleotide (C) in position 44 and 259 novel nucleotide sequence at the 3' end."							
BASE COUNT	64 a	80 c	100 g	65 t					
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
1	aguuuggcaccauugugaucaagacacccguuuucggugugaagcccauuguncugaagggau	60							
51	AGTTGGCGGACCATGCTGTGATCAGAACCGCTTCGGGTGAAGCAGTGTGTGAAGGGGAG	110							
61	acguccuuucguucacuaaccacaaaacccguucgguugugugaaguccugucguugu	120							
111	ACGTCCTTCTGCTCAGTCACACAAAACCGTCTCGGTTGGTGAAGAGTCTGCTGTCT	170							
121	gggaagcgacgucauuaauncggucguugugugugaagccuacagagcuauuugcgcc	180							
171	GGGAAGCAGTCAGTATTAATTCCTGCTGTGTGTGACGCGCTCAGACGACTTGTCCGC	230							
181	ugugacagagcgauagacaaagcgucagcccccguuuuuuuuuuucacagcgagggcaacc	240							
231	TGTGCAGAGCGTAGTACCAAGGGCTGCACCCCGTTTGTTCACAGCGAGGACCAACC	290							
241	ccgcuuuggaauuaaaacu	259							
LOCUS	HGB18973	357 bp	RNA	VRL	19-NOV-1999				
DEFINITION	Hepatitis GB virus B partial 3'UTR region.								
ACCESSION	Y18973								
VERSION	Y18973.1	GI:6018427							
KEYWORDS	3' UTR.								
SOURCE	Hepatitis GB virus B.								
ORGANISM	Hepatitis GB virus B.								
REFERENCE	Vinuesa; ssRNA positive-strand viruses, no DNA stage; Flaviviridae.								
AUTHORS	1 (bases 1 to 357)								
TITLE	Shardelati A., Scarselli E., Tomei L., Kekule A.S. and Traboni C.								
JOURNAL	Identification of a novel sequence at the 3' end of the GB virus B genome								
MEDLINE	J. Virol. 73 (12), 10546-10550 (1999)								
REFERENCE	20027287								
AUTHORS	2 (bases 1 to 357)								
TITLE	Traboni C.								
JOURNAL	Direct Submission								
FEATURES	Submitted (29-APR-1999) C. Traboni, IRBM P. Angeletti, Ist. di Ricerche di Biologia Molecolare, Via Pontina. km. 30.600, 00040 Pomezia, Rome, ITALY								
SOURCE	Location/Qualifiers								
1..357	/organism="Hepatitis GB virus B"								
/specific_host="tamalin"									
/db_xref="taxon:39113"									
1..357	/evidence=experimental								
31..48	/note="poly-U"								
/evidence=experimental									
99..357	/note="3' y region"								
/evidence=experimental									
77 a	88 c	101 g	91 t						





```

repeat_region      /rpt_family="L2"          45623. .46229
                    /rpt_family="L2"          46230. .46277
repeat_region      /rpt_family="MER73-group" 46278. .46504
                    /rpt_family="L2"          46517. .46674
repeat_region      /rpt_family="L1"          46715. .46852
repeat_region      /rpt_family="MER1_type"    47387. .47478
                    /rpt_family="L1"          47824. .48086
repeat_region

Query Match       12.8%; Score 33.2; DB 9; Length 185516;
Best Local Similarity 36.4%; Pred. No. 8.8;
Matches 43; Conservative 22; Mismatches 53; Indels 0; Gaps 0;

QY   60   gacgcccucnucggcgaccacacaacgcgucngggugugaggaaggcucggcugug 119
      ||| : ::| | : | | | | | | | | | | | | | | | | | | | | | | | |
Db 111357 GAGGCGGTTCCTCCCTCCCTCCGCCACAGAATCGGCCGAATAACTTCACTCATGCATGGC 111416
      ||| : ::| | : | | | | | | | | | | | | | | | | | | | | | | | |

QY   120  ugggaagccagcucauuauaacuccgucguqugugugagccuccagcagcuauuuugc 177
      : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 111417 TGAGCAGTGCGTAAGATATTGTCTATGGTGTATGATGTGCACAGAATGACTTTATGGC 111474
      : | : | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 12
LOCUS           AL450993
DEFINITION      Homo sapiens chromosome 1 clone RP11-212D5, *** SEQUENCING IN
ACCESSION       AL450993 AC025357
VERSION         AL450993.7 GI:14625625
KEYWORDS        HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE          human.
ORGANISM        Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 289563)
McLay,K.
Direct Submission
Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 6, 2001 this sequence version replaced gi:14588715.
Draft Sequence Produced by Whitehead Institute/MIT Center for
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba212D5
----- Summary Statistics
Assembly program: XGAP; version 4.5
Sequencing vector: M13; M77815; 44% of reads
Sequencing method: plasmid; L08752; 55% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 265164 bases at least Q40
Consensus quality: 275491 bases at least Q20
Consensus quality: 281753 bases at least Q20
Insert size: 285763; sum-of-contigs
Quality coverage: 3.27x in Q20 bases; sum-of-contigs Quality
coverage: 6.17x in Q20 bases; agarose-fp

** NOTE: This is a 'working draft' sequence. It currently
* consists of 39 contigs. The true order of the pieces
```

```

region      /rpt_family="L2"
            45623. .46229
region      /rpl_family="L2"
            46230. .46277
region      /rpt_family="MER73-group"
            46278. .46504
region      /rpl_family="L2"
            46317. .46674
region      /rpl_family="L1"
            46715. .46852
region      /rpt_family="MR_L_type"
            47387. .47478
region      /rpl_family="L1"
            47824. .48086

ch          12.8%; Score 33.2; DB 9; Length 185516;
Similarity 36.4%; Pred. No. 8;
43; Conservative 22; Mismatches 53; Indels 0; Gaps

AGGCTGGTTTCCTCCCTCCACACAGAATCGCGGAAACTTACTCCTCATCATG 1
cgcgcccucugcgcauccacccacaacacgcgucggugugugagagcucgucgug 1
::: |:| |||| |::| ||| : : : : : : : : : : : : : : : :
ggaagcagcagcaguanuuuuccgucgugugugugugagcgcacagcguuuugc 177
::: |::| ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|
AL450993    289563 bp      DNA           HTG           04-JUL-2000
Homo sapiens chromosome 1 clone RP11-212D5, *** SEQUENCING I
PROGRESS ***, 39 unordered pieces.
AL450993 AC025357
AL450993.7 GI:14625625
HTG: HTGS_PHASE1; HTGS_DRAFT.
human.
Homo sapiens Metazoa: Chordata; Craniata; Vertebrata; Euteleostei
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 289563)
McLay,K.
Direct Submission
Submitted (03-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 6, 2001 this sequence version replaced gi:14588715.
Genome Research, 320 Charles Street,
Cambridge, MA 02141, USA
http://www-seq.wi.mit.edu
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba212nd5
----- Summary Statistics
Assembly program: XGAP4, version 4.5
Sequencing vector: M13; M7815; 44% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 265164 bases at least Q40
Consensus quality: 275491 bases at least Q20
Insert size: 285753; sum-of-contigs
Insert size: 170183; 6.7% error; agarose-fp
Quality coverage: 3.27x in Q20 bases; sum-of-contigs Quality
coverage: 6.17x in Q20 bases; agarose-fp

NOTE: This is a 'working draft' sequence. It currently
consists of 39 contigs. The true order of the pieces

```

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved. 9307: contig of 9307 bp in length

1  
\* 9308 9407: gap of 100 bp  
\* 9408 13091: contig of 3684 bp in length  
\* 13092 13191: gap of 100 bp  
\* 13192 16963: contig of 3772 bp in length  
\* 16964 17063: gap of 100 bp  
\* 17064 28043: contig of 10980 bp in length  
\* 28044 28143: gap of 100 bp  
\* 28144 32213: contig of 4070 bp in length  
\* 32214 32313: gap of 100 bp  
\* 32314 39600: contig of 7287 bp in length  
\* 39601 39700: gap of 100 bp  
\* 39701 41803: contig of 2103 bp in length  
\* 41804 41903: gap of 100 bp  
\* 41904 47480: contig of 5577 bp in length  
\* 47481 47580: gap of 100 bp  
\* 47581 50607: contig of 3027 bp in length  
\* 50608 50707: gap of 100 bp  
\* 50708 65499: contig of 14792 bp in length  
\* 65500 65599: gap of 100 bp  
\* 65600 70500: contig of 4901 bp in length  
\* 70501 70600: gap of 100 bp  
\* 70601 78801: contig of 8201 bp in length  
\* 78802 78901: gap of 100 bp  
\* 78902 88610: contig of 9709 bp in length  
\* 88611 88710: gap of 100 bp  
\* 88711 91276: contig of 2566 bp in length  
\* 91277 91376: gap of 100 bp  
\* 91377 103516: contig of 12140 bp in length  
\* 103517 103616: gap of 100 bp  
\* 103617 106242: contig of 2626 bp in length  
\* 106243 106342: gap of 100 bp  
\* 106343 114051: contig of 7709 bp in length  
\* 114052 114151: gap of 100 bp  
\* 114152 128614: contig of 14463 bp in length  
\* 128615 128714: gap of 100 bp  
\* 128715 134258: contig of 5544 bp in length  
\* 134259 134358: gap of 100 bp  
\* 134359 139476: contig of 5118 bp in length  
\* 139477 139576: gap of 100 bp  
\* 139577 149666: contig of 10290 bp in length  
\* 149667 149966: gap of 100 bp  
\* 149967 159641: contig of 9675 bp in length  
\* 159642 159741: gap of 100 bp  
\* 159742 173858: contig of 14117 bp in length  
\* 173859 173958: gap of 100 bp  
\* 173959 181673: contig of 7715 bp in length  
\* 181674 181773: gap of 100 bp  
\* 181774 186169: contig of 4396 bp in length  
\* 186170 186269: gap of 100 bp  
\* 186270 201406: contig of 15137 bp in length  
\* 201407 201506: gap of 100 bp  
\* 201507 203507: contig of 2001 bp in length  
\* 203508 203607: gap of 100 bp  
\* 203608 227405: contig of 23798 bp in length  
\* 227406 227505: gap of 100 bp  
\* 227506 230659: contig of 3154 bp in length  
\* 230660 230759: gap of 100 bp  
\* 230760 233331: contig of 2572 bp in length  
\* 233332 233431: gap of 100 bp  
\* 233432 236751: contig of 3320 bp in length  
\* 236752 236851: gap of 100 bp  
\* 236852 242617: contig of 5766 bp in length  
\* 242618 242717: gap of 100 bp  
\* 242718 245181: contig of 2464 bp in length  
\* 245182 245281: gap of 100 bp  
\* 245282 257886: contig of 12605 bp in length

## FEATURES

## source

\* 257887 257986: gap of 100 bp  
\* 257987 265713: contig of 7727 bp in length  
\* 265714 265813: gap of 100 bp  
\* 265814 271565: contig of 5752 bp in length  
\* 271566 271665: gap of 100 bp  
\* 271666 276811: contig of 5146 bp in length  
\* 276812 276911: gap of 100 bp  
\* 276912 281191: contig of 4280 bp in length  
\* 281192 281291: gap of 100 bp  
\* 281292 289563: contig of 8272 bp in length.  
Location/Qualifiers  
1. 289563  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone\_RP1="212D5"  
/clone\_lib="RPCT-11.1"  
1. 9307

## misc\_feature

/note="assembly-fragment:00169  
fragment\_chain:1"

## misc\_feature

/note="assembly-fragment:00741  
fragment\_chain:1"

## misc\_feature

/note="assembly-fragment:01915  
fragment\_chain:1"

## misc\_feature

/note="assembly-fragment:00207  
fragment\_chain:2"

## misc\_feature

/note="assembly-fragment:00775  
fragment\_chain:2"

## misc\_feature

/note="assembly-fragment:00306  
fragment\_chain:2"

## misc\_feature

/note="assembly-fragment:00535  
fragment\_chain:3"

## misc\_feature

/note="assembly-fragment:00988  
fragment\_chain:3"

## misc\_feature

/note="assembly-fragment:01502  
fragment\_chain:4"

## misc\_feature

/note="assembly-fragment:00022  
fragment\_chain:4"

## misc\_feature

/note="assembly-fragment:01705  
fragment\_chain:5"

## misc\_feature

/note="assembly-fragment:02128  
fragment\_chain:5"

## misc\_feature

/note="assembly-fragment:01837  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00450  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00080"  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00130"  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00446"  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00448"  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00458"  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:00731"  
fragment\_chain:6"

## misc\_feature

/note="assembly-fragment:01002"  
fragment\_chain:6"



repeat_region	3338..3465	/rpt_family="LINE/L2"
repeat_region	3941..4089	/rpt_family="LINE/L2"
repeat_region	4332..4410	/rpt_family="LINE/L2"
repeat_region	4417..4560	/rpt_family="LINE/L2"
repeat_region	5509..5584	/rpt_family="SINE/B2"
misc_feature	5600..5727	/note="Region containing CAAA)n"
repeat_region	5807..5851	/rpt_family="L1R/MALR"
repeat_region	5820..6050	/rpt_family="L1R/MALR"
repeat_region	6051..6243	/rpt_family="SINE/B2"
repeat_region	6244..6407	/rpt_family="L1R/MALR"
misc_feature	7282..7305	/note="identical to 7406..7429"
misc_feature	7406..7429	/note="identical to 7282..7305"
misc_feature	7508..7535	/note="identical to 7554..7581"
misc_feature	7554..7581	/note="identical to 7508..7535"
repeat_region	9065..9256	/rpt_family="L1R"
repeat_region	9257..9355	/rpt_family="7C)n"
repeat_region	9356..9564	/rpt_family="L1R"
repeat_region	9697..9929	/rpt_family="LINE/L1"
repeat_region	10016..10067	/rpt_family="(CCA)n"
misc_feature	10540..10803	/note="Region conserved with human (AC005294)"
repeat_region	10808..10885	/rpt_family="AT-rich"
repeat_region	11225..11637	/rpt_family="L1R/MALR"
repeat_region	12879..13008	/rpt_family="SINE/Alu"
repeat_region	12912..13019	/rpt_family="SINE/B4"
misc_feature	13039..13068	/note="identical to 13067..13096"
misc_feature	13067..13096	/note="identical to 13039..13068"
repeat_region	13193..13240	/rpt_family="L1R/Retroviral"
repeat_region	13241..13312	/rpt_family="(7G)n"
repeat_region	13313..13402	/rpt_family="L1R/Retroviral"
repeat_region	14077..14181	/rpt_family="LINE/L1"
repeat_region	14182..14223	/rpt_family="(CTA)n"
repeat_region	14224..14404	/rpt_family="LINE/L1"
misc_feature	16814..16975	/note="Region conserved with human (AC005294)"
repeat_region	18915..19059	/rpt_family="SINE/Alu"
repeat_region	19063..19101	/rpt_family="(CAA)n"
repeat_region	19317..19752	/rpt_family="LINE/L1"
repeat_region	20508..20552	







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 14:28:12 ; Search time 166.92 Seconds  
(without alignments)  
1330.261 Million cell updates/sec

Title: US-09-587-653-1

Perfect score: 259  
Sequence: 1 aguugggcgaccgaugugga.....cccccuuggaauaaaaacu 259

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries:

Database :

1: N\_Geneseq\_1101.\*  
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	259	100.0	9399	22 AAF23484	GBV-B virus genome
2	257.4	99.4	259	22 AAC91998	GBV-B 3'X DNA sequ
3	257.4	99.4	22	22 AAC92010	GBV-B 3'X RNA sequ
4	245.4	94.7	258	22 AAC91996	GBV-B 3'X RNA sequ
5	82	31.7	82	22 AAC92012	Human transporter
6	32	12.4	3179	22 AAD09560	Human secreted pro
7	30.4	11.7	522	21 AAC81741	Human secreted pro
8	30.4	11.7	2342	21 AAA93110	Human LOBO homolog
9	30	11.6	49999	20 AAZ23900	Human gene for col
10	29.8	11.5	24183	22 AAS21771	Human RNA-binding
11	29.4	11.4	1933	22 AAZ20379	

C 12	29.4	11.4	1993	22 AAH15309	Human cDNA sequenc
C 13	29.4	11.4	2092	21 AAA49680	Human quaking spli
C 14	29.4	11.4	2472	21 AAA49681	Human quaking spli
C 15	29	11.2	775	21 AAC39642	Arabidopsis thalia
C 16	28.8	11.1	400	22 AAH97854	Murine 7-transmemb
C 17	28.6	11.0	319608	21 AAH51601	Human chromosome 1
C 18	28.4	11.0	319608	22 AAS09301	Human schizophreni
C 19	28.4	11.0	531	22 AAH97850	Murine 7-transmemb
C 20	28.4	11.0	560	22 AAH97844	Murine 7-transmemb
C 21	28.2	10.9	1548	22 AAF93854	Human cDNA encodin
C 22	28.2	10.9	6623	19 AAV19051	Mitochondrial gene
C 23	28.2	10.9	6623	19 AAV19053	Mitochondrial gene
C 24	28	10.8	1723	13 AAQ28737	Human osteogenic p
C 25	28	10.8	1723	14 AAQ53155	Sequence encoding
C 26	28	10.8	1723	14 AAQ38946	Human osteogenic p
C 27	28	10.8	1723	14 AAQ38735	Human OP-2. Homo
C 28	28	10.8	1723	14 AAQ38859	Morphogen hOP2 cod
C 29	28	10.8	1723	15 AAQ56200	hOP2 cDNA. Homo s
C 30	28	10.8	1723	15 AAQ56233	hOP2 cDNA. Homo s
C 31	28	10.8	1723	15 AAQ57917	Human osteogenic p
C 32	28	10.8	1723	15 AAQ67313	Human OP-2. Homo
C 33	28	10.8	1723	15 AAQ45118	Human OP-2. Homo
C 34	28	10.8	1723	15 AAQ65393	Osteogenic protein
C 35	28	10.8	1723	15 AAQ45164	Osteogenic protein
C 36	28	10.8	1723	15 AAQ71426	hOP2-PF prepro for
C 37	28	10.8	1723	15 AAQ72706	hOP2 cDNA. Homo
C 38	28	10.8	1723	17 AAT33443	Human osteogenic p
C 39	28	10.8	1723	17 AAT02600	hOP-2 cDNA. Homo
C 40	28	10.8	1723	18 AAV10347	Human OP-2. Homo
C 41	28	10.8	1723	18 AAT97881	Human osteogenic p
C 42	28	10.8	1723	18 AAT73208	Human osteogenic p
C 43	28	10.8	1723	19 AAV32585	Human osteogenic p
C 44	28	10.8	1723	19 AAV19535	Human osteogenic p
C 45	28	10.8	1723	22 AAC89690	Human tissue morph

#### ALIGNMENTS

RESULT 1	AAAF23484	standard; DNA; 9399 BP.
ID	AAAF23484	
XX	AAAF23484	
AC	21-MAR-2001	(first entry)
XX		
DE	GBV-B virus genome.	
XX		
KW	GBV-B; hepatitis C virus; HCV; vaccine; ds.	
XX		
OS	GB virus B.	
XX		
XX	WO200075337-A1.	
XX		
PD	14-DEC-2000.	
XX		
PF	02-JUN-2000; 2000WO-US15293.	
XX		
PR	04-JUN-1999; 99US-0137694.	
XX		
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Bukh J, Yanagi M, Emerson SU, Purcell RH;	
XX		
DR	WPI; 2001-091214/10.	
XX		
PT	New infectious nucleic acids of the GB virus-B clone, useful for	
PT	indirectly studying the molecular properties of hepatitis C virus (HCV)	
XX	and in developing vaccines and therapeutics for HCV	
PS	Claim 3; Page 60-63; 96pp; English.	

CC The present invention relates to GB virus-B. The nucleic acid molecules  
CC of the invention are useful for indirectly studying the molecular  
CC properties of hepatitis C virus (HCV). The infectious nucleic acid  
CC sequence of the GB virus-B clone and the HCV/GBV-B chimeras may be used  
CC in the development of vaccines and therapeutics for HCV.

AA Sequence 9399 BP; 2125 A; 2326 C; 2430 G; 2518 T; 0 other;  
SQ

Query Match	100.0%	Score 259;	DB 22,	Length 9399;
Best Local Similarity	76.1%;	pred.No. 1.4e-78;		
Matches 197; Conservative	62;	Mismatches	0;	Indels 0; Gaps 0;

OY	1	agunungcagacagugugugugacacccgubuuocgugugagacagugucubaaagugag	60
Db	9141	agttctgagacacatgvgvgatccagaaaccgcttccggtgtagaacacatgctctgaagggagtg	9200
OY	61	acgucuccuuuugugucubauucacacaaacccgucucgvgugugvgagagagucucugugucug	120
Db	9201	acgtccctctctgctcattccacaaacacccgtctcgtggtgvgtagagagctcctgtgtgtgt	9260
OY	121	ggggaagcagucaguanauuucccgucugugugugagacgucucacagagacuguuugucacgc	180
Db	9261	ggggaagcagctcatatcatattccctcgtctgtgtgtgtgtagcctccacagacgtattctgtccgc	9320
OY	181	ugugcagagcagugugaaccaagggucgucaccccguguuuuuguuuguuucacagcgggggacaccc	240
Db	9321	tgtgtcagaagcgtgtgtacacaaaggctgtcaccccggtttgtgttccaaagcggagggacaccc	9380
OY	241	ccgcuguggaauaaaaaacu	259
Db	9381	ccgctgtgaattaaacat	9399

## RESULT 2

AAC91998 standard; DNA; 259 BP.

AAC91998;

DT 20-MAR-2001 (first entry)

GBV-B 3' X DNA sequence.

GBV-B; hepatitis GB viral-B; hepatitis C virus activity; ds.

05 Hepatitis virus.

PN WO200073466-A1.

PD 07-DEC-2000

PF 22-MAY-2000; 2000WO-EP04622.

PR 27-MAY-1999; 99GB-0012432.

PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.

PI Traboni C;

DR WPI; 2001-061544/07.

Novel GB viral sequence useful in in vivo assays to identify agents  
PT which modulate hepatitis C virus activity -

PS Claim 1; page 14; 76pp; English.

CC The present sequence is an isolated hepatitis GB viral-B (GBV-B)  
CC polynucleotide. The hepatitis GB agent was first discovered by the  
CC inoculation of tamarins with serum from a patient, whose initials were  
CC GB, affected by acute hepatitis. The serum induced hepatitis in all  
CC inoculated tamarins. The present sequence comprises the 3' termini of  
CC GBV-B, and confers infectivity in tamarins on otherwise non-infective

CC GBV-B genome. This sequence is useful in in vivo assays to identify agents which modulate hepatitis C virus (HCV) activity.

Sequence 259 BP; 53 A; 64 C; 81 G; 61 T; 0 other;

Query match	99.48; Score 257.4; DB 22; Length 259;
-------------	--

[illegible]

### RESULT 3

ID AAC92010 standard; DNA; 309 BP.

AC AAC92010;

DT 20-MAR-2001 (first entry)

GBV-B 3'X sequence and 3'UTR region.

GBV-B; hepatitis GB viral-B; hepatitis C virus activity; ds.

OS Hepatitis virus.

PN WO200073466-A1.

PD 07-DEC-2000

PF 22-MAY-2000; 2000WO-EP04622.

PR 27-MAY-1999; 99GB-0012432.

PA (RICE-) IST RICERCHE BIOL MOLECOLARE ANGELETTI.

PI Traboni C

DR WPI; 2001-061544/07.

Novel GB viral sequence useful in in vivo assays to identify agents  
PT which modulate hepatitis C virus activity -

PS Example 2; Fig 3; 76pp; English.

The present invention relates to a hepatitis GB viral-B (GBV-B) polynucleotide (see AAC91998 and AAC91996). The hepatitis GB agent was first discovered by the inoculation of tamars with serum from a patient, whose initials were GB, affected by acute hepatitis. The serum induced hepatitis in all inoculated tamars. The present sequence comprises the sequence in AAC91996 (the 3' terminus of GBV-B, plus the 3' UTR. This sequence confers infectivity in tamars on otherwise non-infective GBV-B genome. This sequence is useful in vivo assays to identify agents which modulate hepatitis C virus (HCV) activity.

Query Match	99.4%	Score 257.4	DB 22	Length 309
Best Local Similarity	76.1%	Pred. No. 1.4e-78		
Matches 197	Conservative 61	Mismatches 1	Indels 0	Gaps 0
OY	1	aguuuuggcagcaugauguagcaagaccguuuuugggugaaagccauguguaagggagug	60	
DB	51	agtttgcagccatgtgtgatacagaacgtttcttggtgaaagccatgtgtgaaagggatg	110	
OY	61	acguccuuucugugucuaucacacaaaacccugucggugugagagagucucuguguguu	120	
DB	111	agctccctcttgcctatccacaaaaacccgtctgggtgtgtgtgaggaagtcctggtgtgt	170	
OY	121	gggaagcagucaguanauauuuccugucugugugugagcgcucacagcaguanuugccgc	180	
DB	171	gggaagcagctcagtaataatccctgcgtgtgtgtgtagcagccctcagcagctactgtccgc	230	
OY	181	uuugcagcagcuguauguaacaaaggcugaccgccguuuuuuuuucacagcggaaggcaacc	240	
DB	231	tggcagagcgtatgataccaaaggcugaccccggttttcttccaaagcgagggcaacc	290	
OY	241	ccgcuugaguanuuaaacu 259		
DB	291	ccgcttgatataaact 309		
RESULT 4				
ID	AAC91996			
AC	AAC91996 standard; RNA; 258 BP.			
AC	AAC91996;			
-DT	20-MAR-2001 (first entry)			
DE	GBV-B 3'X RNA sequence #1.			
XX	GBV-B; hepatitis GB viral-B; hepatitis C virus activity; ds.			
XX	Hepatitis virus.			
OS	WO200073466-A1.			
XX	07-DEC-2000.			
XX	22-MAY-2000; 2000WO-EP04622.			
XX	27-MAY-1999; <b>99GB-00124321</b>			
PA	(RICE-) IST RICERCHE BIOL MOLECULAIRE ANGELETTI.			
XX	Traboni C;			
PI	WPI; 2001-061544/07.			
DR				
XX				
PT	Novel GB viral sequence useful in in vivo assays to identify agents			
XX	which modulate hepatitis C virus activity			
XX	Claim 1; Page 62; 76pp; English.			
CC	The present sequence is an isolated hepatitis GB viral-B (GBV-B)			
CC	polynucleotide. The hepatitis GB agent was first discovered by the			
CC	inoculation of tamarins with serum from a patient, whose initials were			
CC	GB, affected by acute hepatitis. The serum induced hepatitis in all			
CC	inoculated tamarins. The present sequence comprises the 3' terminus of			
CC	GBV-B, and confers infectivity in tamarins on otherwise non-infective			
CC	GBV-B genome. This sequence is useful in in vivo assays to identify			
CC	agents which modulate hepatitis C virus (HCV) activity.			
XX				
XX	Sequence 258 BP; 53 A; 64 C; 80 G; 61 U; 0 other;			

Query Match	94.7%:	Score 245.4;	DB 22;	Length 258;
Best Local Similarity	99.2%:	Pred. No. 1.7e-74;		
Matches 257;	Conservative 0;	Mismatches 1;	Indels 1;	Gaps 1;
OY	1	aguuugcgaccaauggaucaagaaccgguucggugugaagccaaugucuaaagggaug	60	
Db	1	aguuuugcgaccaugugugaucaagaaccgguucggugugaagccaaugucuaaagggaug	60	
OY	61	acguucccuucuggcucaaccacaataaacccgucucggugugaggaaguccugcuguu	120	
Db	61	acguucccuucuggcucaaccacaataaacccgucucggugugaggaaguccugcuguu	119	
OY	121	gggaagcgacgucaaguaaaucccugucugugugugacgcacacgaauuuugccgc	180	
Db	120	gggaagcgacgucaaguaaaucccugucugugugugacgcacacgaauuuugccgc	179	
OY	181	ugugcagagcguaagaccacgaagcgcccccgguuuuuuguccaagcggaaggcaacc	240	
Db	180	ugugcagagcguaagaccacgaagcgcccccgguuuuuuguccaagcggaaggcaacc	239	
OY	241	ccgcuuuggaauuaaacaacu	259	
Db	240	ccgcuuuggaauuaaacaacu	258	
RESULT	5			
AAC92012				
ID	AAC92012	standard; RNA; 82 BP.		
XX	AAC92012;			
XX	20-MAR-2001	(first entry)		
XX	GBV-B 3'X RNA sequence #2.			
XX	GBV-B; hepatitis GB viral-B; ds.			
XX	Hepatitis virus.			
XX	WO200073466-A1.			
XX	07-DEC-2000.			
XX	22-MAY-2000;	2000WO-EP04622.		
XX	27-MAY-1999;	99GB-0012432.		
PA	(RICE-) IST RICHECHE BIOL MOLECULAIRE ANGELETTI.			
PI	Traoonl C;			
DR	WPI: 2001-061544/07.			
PT	Novel GB viral sequence useful in vivo assays to identify agents			
PS	which modulate hepatitis C virus activity	-		
CC	Disclosure; Fig 4; 76pp: English.			
CC	The present sequence is an isolated hepatitis GB viral-B (GBV-B)			
CC	polynucleotide. The hepatitis GB agent was first discovered by the			
CC	inoculation of tamarins with serum from a patient, whose initials were			
CC	GB, affected by acute hepatitis. The serum induced hepatitis in all			
CC	inoculated tamarins. The present sequence comprises the 3' terminus of			
CC	GBV-B.			
SQ	Sequence 82 BP; 19 A; 23 G; 23 G; 17 U; 0 other;			
Query Match	31.7%:	Score 82;	DB 22;	Length 82;
Best Local Similarity	100.0%:	Pred. No. 1.2e-18;		
Matches 82;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;







```

XX XX WO9950284-A2.
XX PN
XX PD 07-OCT-1999.
XX PF 26-MAR-1999; 99WO-EP02055.
XX PR 27-MAR-1998; 98DE-1013799.
XX PA (ROSE/) ROSENTHAL A.
XX PI Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX DR WPI, 1999-601320/51.
XX PS
XX PT Nucleic acids encoding proteins which influence bone development,
XX PR useful for treating and studying bone disorders -
XX PS
XX PS Example 3: Page 244-271; 391pp; German.
XX PS
XX CC This invention describes novel nucleic acids (i; designated LOBO (long
XX CC bones)) encoding proteins influencing bone development in mammals. The
XX CC proteins of the invention reduce and/or inactivate bone extension (i.e.
XX CC development), with exception of the skull and have osteopathic activity.
XX CC The nucleic acid molecules, proteins and antibodies can be used in
XX CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX CC and nucleic acid molecules, etc. are useful for production of transgenic
XX CC animals, especially a transgenic mouse for the study of diseases
XX CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX CC achondroplasia. This sequence encodes a human LOBO protein described
XX CC in the method of the invention.
XX SQ Sequence 49999 BP; 11432 A; 13017 C; 12505 G; 13045 T; 0 other;

Query Match 11.6%; Score 30; DB 20; Length 49999;
Best Local Similarity 39.8%; Pred. No. 7.9;
Matches 47; Conservative 16; Mismatches 55; Indels 0; Gaps 0;

QY 14 uggugauagacacgucgugugugagacgucgugaggggagucgucucucg 73
DB 24459 tgggtgaltcagcagcacttcagttgggacattgtcagtagtagagatcctagtgatgc 24518
QY 74 cucuaccacaaaacacgucgugugugagagucgucgucgucgucgucgucgucg 131
DB 24519 cctcttcgtcgctgcgacttgagaggggacagcagcagcggtgtctcgtgatgaagtc 24576

RESULT 10
AAS21771/C
ID AAS21771 standard; DNA; 24183 BP.
XX
XX AAS21771;
XX AC
XX XX 24-OCT-2001 (first entry)
XX DT
XX XX
XX DE Human gene for collagen COL9A1.
XX
XX KM Human; collagen; COL1A1; COL1A2; COL9A1; COL9A2; COL9A3; ds;
XX KM osteoporosis; multiple epiphyseal dysplasia; osteogenesis imperfecta;
XX KM shortness of stature; low bone density; gene therapy.
XX XX
XX OS Homo sapiens.
XX PN US6265157-B1.
XX PD 24-JUL-2001.
XX PF 03-OCT-1997; 97US-0943731.
XX PR 03-DEC-1991; 91US-0803628.
XX PR 13-MAR-1994; 94US-0212322.
XX

```

```

PA (UYAL-) UNIV ALLEGHENY HEALTH SCI.
PA (UYAL-) UNIV ALLEGHENY HEALTH SCI.
PA (UYAL-) UNIV ALLEGHENY HEALTH SCI.
XX (YOU-) UNIV OULU.
XX
XX Prockop DJ, Spottila LD, Deltas CD, Sereida L, Westerhausen Larson A;
XX PI Pack M, Collige A, Early J, Koerkhoe J, Ala-kokko L, Annunen S;
XX PI Pihlajamaa T, Vuoristo M, Paasilta P;
XX DR WPI, 2001-432201/46.
XX
XX PT Detecting collagen gene alteration, useful for diagnosing osteoporosis,
XX PR multiple epiphyseal dysplasia, osteogenesis imperfecta, shortness of
XX PR stature and low bone density in humans -
XX PS
XX PS Example 5; Fig 10A; 617pp; English.
XX
XX CC The invention relates to detecting a collagen gene alteration associated
XX CC with a pathological condition in a human subject by obtaining from the
XX CC subject a sample nucleic acid containing a portion of at least 15
XX CC consecutive nucleotides of the segment of the COL1A1 gene extending in
XX CC the 5' to 3' direction from 78 nucleotides of intron 27 located adjacent
XX CC exon 28 through the 3' end of intron 51, where the portion contains an
XX CC intronic nucleotide and a first and second site, determining the sequence
XX CC of the portion and comparing the sequence of the portion with the
XX CC corresponding consensus sequence of the COL1A1 gene where a difference
XX CC between the sequence of the portion and the consensus sequence indicates
XX CC the presence of the collagen alteration in the subject. The method is
XX CC used for detecting abnormalities in a COL1 or COL2 gene is useful for
XX CC determining whether a subject is afflicted with pathological conditions
XX CC associated with an altered collagen gene such as osteoporosis, multiple
XX CC epiphyseal dysplasia, osteogenesis imperfecta, shortness of stature and
XX CC low bone density. Identification of an abnormality in a collagen gene is
XX CC also useful for designing a therapeutic nucleotide or gene therapy agent
XX CC which can be administered to the subject to correct or alleviate the
XX CC abnormality. The method is useful for detecting mutations in both the
XX CC coding and non-coding sequences of any of the COL1 or COL2 genes.
XX CC Therefore the method can be used to detect collagen gene alterations
XX CC which affect either the primary sequence of a collagen protein chain, of
XX CC splicing of the mRNA encoding such chains or regulation of expression of
XX CC the genes encoding such chains. The present sequence is a collagen gene
XX CC of the invention.
XX SQ Sequence 24183 BP; 6756 A; 4685 C; 5006 G; 7607 T; 129 other;

Query Match 11.5%; Score 29.8; DB 22; Length 24183;
Best Local Similarity 38.1%; Pred. No. 7.1;
Matches 43; Conservative 18; Mismatches 52; Indels 0; Gaps 0;

QY 120 uggagagcagucaguanuauuuccgucgugugugagcgcucacgacgucgucg 179
DB 4413 ttgcagagtgatgcagcagcagcagcagcagcagcagcagcagcagcagcagc 4354
QY 180 cugugcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 232
DB 4353 ctggccgcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 4301

RESULT 11
AAH20379/C
ID AAH20379 standard; CDNA; 1933 BP.
XX
XX AAH20379;
XX AC
XX XX 03-AUG-2001 (first entry)
XX DT
XX XX
XX DE Human RNA-binding protein 33 CDNA.
XX
XX KM Human; RNA-binding protein 33; tumour; haemopathy; HIV infection;
XX KM immunological disease; inflammatory disease; ss.
XX OS Homo sapiens.
XX

```

[illegible]

```

RESULT 13
ID AAA49680/C
XX AAA49680 standard; DNA; 2092 BP.
AC AAA49680;
XX
XX 25-SEP-2000 (first entry)
DT
XX Human quaking splice variant 5 (OK5).
XX
XX Quaking splice variant 5; OK5 gene; human;
XX differential gene expression; movement disorder; motor pathology;
XX dystonia; dyskinesia; therapy; diagnosis; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 28..1053
FT /*tag= a
XX
XX WO200037685-A2.
XX
XX 29-JUN-2000.
XX
XX 21-DEC-1999; 99WO-US30727.
XX
XX 21-DEC-1998; 98US-0113127.
XX 20-DEC-1999; 99US-0468253.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg B;
XX
XX WPI: 2000-442696/38.
XX
XX P-PSDB; AAY95336.
XX
XX Identifying psychotropic agents useful in the treatment of e.g.
XX schizophrenia comprises comparing the expression patterns of HALO genes
XX in test and control cell populations -
XX
XX Disclosure; Fig 1; 93pp; English.
XX
XX The present sequence of the human quaking splice variant 5 (OK5),
XX a homologue of the mouse quaking type 1 gene, which is a member of
XX the STAR (signal transduction and activator of RNA) class of
XX proteins. Human OK5 was isolated following the identification of a
XX rat homologue (see AAA49679) of mouse quaking. The rat gene was
XX shown to be differentially expressed in response to the neuroleptic
XX haloperidol. The human OK5 gene is useful as a marker for the
XX onset of tardive dyskinesia/dystonias in human subjects taking
XX neuroleptics. Determination of its level of expression, and
XX manipulation of its level of expression, are used in methods of
XX diagnosing or determining susceptibility to a movement disorder,
XX and in methods of preventing or delaying the onset of a motor
XX pathology, especially dystonia.
XX
XX Sequence 2092 BP; 683 A; 433 C; 449 G; 527 T; 0 other;
XX
Query Match 11.4%; Score 29.4; DB 21; Length 2092;
Best Local Similarity 38.9%; Pred. No. 4;
Matches 79; Conservative 23; Mismatches 96; Indels 5; Gaps 1;
XX
XX 35 ggugaagccaugucugaggggaugagcuccuucugcucacacaaacgucuc 94
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX 900 GGGGACTCATGTAGCTTAACCAAGCTTCGGCCCTGGAGGAATATTGCACG 841
XX
XX 95 ggugugugagugagucucugugugugaggaagacagucaguanauucguc- 149
XX |::|||:||||:||||:||||:||||:||||:||||:||||:||||:
XX 840 AGTTGGGTAGAGAGTTCCCTTTGGCATGACACGGCTCTATTTGCTATCAAGGCA 781
XX
XX 150 ugugugagcgcacagcagcaguanuuugucgucugucagagcugagucacaaagcgucgac 209
XX :|||: ||| | ||| | ||| | ||| | ||| | ||| | |||

```

```

DB 780 TATGTAGGGCCAGCTGGCGTAGGACTACGACGGACGACTGTGGAGAACCGGCGCAG 721
OY 210 cccgguuuuuugucacagcgag 232
||| |: : | | | |
DB 720 CCCAGTATGATCTCTTGGACAG 698

RESULT 14
ID AAA49681/C
XX AAA49681 standard; DNA; 2472 BP.
AC AAA49681;
XX
XX 25-SEP-2000 (first entry)
DT
XX Human quaking splice variant 7 (OK7).
XX
XX Quaking splice variant 7; OK7 gene; human;
XX differential gene expression; movement disorder; motor pathology;
XX dystonia; dyskinesia; therapy; diagnosis; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH CDS 6..983
FT /*tag= a
XX
XX WO200037685-A2.
XX
XX 29-JUN-2000.
XX
XX 21-DEC-1999; 99WO-US30727.
XX
XX 21-DEC-1998; 98US-0113127.
XX 20-DEC-1999; 99US-0468253.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Gould-Rothberg B;
XX
XX WPI: 2000-442696/38.
XX
XX P-PSDB; AAY95335.
XX
XX Identifying psychotropic agents useful in the treatment of e.g.
XX schizophrenia comprises comparing the expression patterns of HALO genes
XX in test and control cell populations -
XX
XX Disclosure; Fig 2; 93pp; English.
XX
XX The present sequence of the human quaking splice variant 7 (OK7),
XX a homologue of the mouse quaking type 1 gene, which is a member of
XX the STAR (signal transduction and activator of RNA) class of
XX proteins. Human OK7 was isolated following the identification of a
XX rat homologue (see AAA49679) of mouse quaking. The rat gene was
XX shown to be differentially expressed in response to the neuroleptic
XX haloperidol. The human OK5 gene is useful as a marker for the
XX onset of tardive dyskinesia/dystonias in human subjects taking
XX neuroleptics. Determination of its level of expression, and
XX manipulation of its level of expression, are used in methods of
XX diagnosing or determining susceptibility to a movement disorder,
XX and in methods of preventing or delaying the onset of a motor
XX pathology, especially dystonia.
XX
XX Sequence 2472 BP; 754 A; 490 C; 500 G; 728 T; 0 other;
XX
Query Match 11.4%; Score 29.4; DB 21; Length 2472;
Best Local Similarity 38.9%; Pred. No. 4.3;
Matches 79; Conservative 23; Mismatches 96; Indels 5; Gaps 1;
XX
XX 35 ggugaagccaugucugaggggaugagcuccuucugcucacacaaacgucuc 94
XX |||||:||||:||||:||||:||||:||||:||||:||||:||||:
XX 878 GGGGACTCATGTAGCTTAACCAAGCTTCGGCCCTGGAGGAATATTGCACG 819

```















```

US-07-841,646-28
: Sequence 28, Application US/07841646
: Patent No. 5266683
:
: GENERAL INFORMATION:
:
: APPLICANT: OPPERMAN, HERMANN
:
: APPLICANT: KUBERANK, ENGIN
:
: APPLICANT: KUBERANKAPATHI, THIANGAVEL
:
: APPLICANT: RUEGER, DAVID C.
:
: APPLICANT: PANG, ROY H. L.
:
: TITLE OF INVENTION: OSTEOGENIC DEVICES
:
: NUMBER OF SEQUENCES: 33
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: TESTA, HURWITZ & THIBEAULT
:
: STREET: 53 STATE STREET
:
: CITY: BOSTON
:
: STATE: MASSACHUSETTS
:
: COUNTRY: U.S.A.
:
: ZIP: 02109
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Floppy disk
:
: COMPUTER: IBM PC compatible
:
: OPERATING SYSTEM: PC-DOS/MS-DOS
:
: SOFTWARE: Patent Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/07/841,646
:
: FILING DATE: 19920221
:
: CLASSIFICATION: 530
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 810,560
:
: FILING DATE: 20-DEC-1991
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 827,052
:
: FILING DATE: 28-JAN-1992
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 660,162
:
: FILING DATE: 22-FEB-1991
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 621,988
:
: FILING DATE: 04-DEC-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 621,849
:
: FILING DATE: 04-DEC-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 616,374
:
: FILING DATE: 21-NOV-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 600,024
:
: FILING DATE: 18-OCT-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 599,543
:
: FILING DATE: 18-OCT-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 579,865
:
: FILING DATE: 07-SEP-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 569,920
:
: FILING DATE: 20-AUG-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 483,913
:
: FILING DATE: 22-FEB-1990
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 422,613
:
: FILING DATE: 17-OCT-1989
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 315,342
:
: FILING DATE: 23-FEB-1989
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 232,630
:
: FILING DATE: 15-AUG-1988
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: US 179,460
:
: FILING DATE: 08-APR-1988
:
: ATTORNEY/AGENT INFORMATION:

```

```

NAME: PITCHER, EDEMUND R.
REGISTRATION NUMBER: 27, 829
REFERENCE/DOCKET NUMBER: CRP-001Cp6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1723 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: HIPPOCAMPUS
FEATURE:
NAME/KEY: CDS
LOCATION: 490..1696
OTHER INFORMATION: /function="OSTROGEN
OTHER INFORMATION: /product="hnp2-p"
OTHER INFORMATION: /note="hnp2 (cDNA)"

```

Query Match	10.8%;	Score 28;	DB 1;	Length 1723;
Best Local Similarity	47.0%;	Pred. No. 2.2;		
Matches	47;	Conservative	8;	Mismatches 45;
				Indels 0;
				Gaps 0;

[illegible]

RESULT 5  
US-07-901-703-10  
: Sequence 10, Application US/07901703  
Patent No. 5344654  
GENERAL INFORMATION:  
APPLICANT: RUEGER, DAVID C  
APPLICANT: KUBERASAMPATH, THANGAVEL  
APPLICANT: OPPERMAN, HERMAN  
APPLICANT: OZAKYANAK, GENIN  
TITLE OF INVENTION: PROSTHETIC DEVICES HAVING ENHANCED  
TITLE OF INVENTION: OSTEOGENIC PROPERTIES  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TESTA, HURWITZ & TRIBAULT  
STREET: EXCHANGE PLACE, 53 STATE STREET  
CITY: BOSTON  
STATE: MA  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/901,703  
FILING DATE: 19920616  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: PITCHER ESO, EDMUND R  
REGISTRATION NUMBER: 27,829  
REFERENCE/DOCKET NUMBER: STK-057  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/248-7000  
INFORMATION FOR SEQ ID NO: 10:

```
SEQUENCE CHARACTERISTICS:
LENGTH: 1723 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: HIPPOCAMPUS
FEATURE:
NAME/KEY: CDS
LOCATION: 490..1696
OTHER INFORMATION: /function="OSTEOGENIC PROTEIN"
OTHER INFORMATION: /product="hop2-p"
OTHER INFORMATION: /note="hop2 (cDNA)"
US-07-901-703-10

Query Match      10.8%; Score 28; DB 1; Length 1723;
Best Local Similarity 47.0%; Pred. No. 2.2;
Matches 47; Conservative 8; Mismatches 45; Indels 0; Gaps 0;

QY 1 aguuuugcgacgaugugaucaagaacgguuucgguguaagccauugucugaagggaug 60
    ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 808 AGCTTCGTTAAACATGATGTGAGAGAGACCGCTGCGCCCTGGGCCACACGAGAGCCCATTTGGAAG 867

QY 61 acguccuucucugucuaucacaaacccgucucggugug 100
    | | : | | | | | | | | | | | | | | | | | | |
DB 866 GAGTTCGCTTGACCTGACCCAGATCCCGGCTGGGAGG 907

RESULT 6
US-08-147-023-28
; Sequence 28, Application US/08147023
; Patent No. 5468845
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: OSTEOGENIC DEVICES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/147,023
; FILING DATE: 21-FEB-1992
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 810,560
; FILING DATE: 20-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 827,052
; FILING DATE: 28-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 660,162
; FILING DATE: 22-FEB-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,988
; FILING DATE: 04-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 621,849
```

```
FILING DATE: 04-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 616,374
FILING DATE: 21-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 600,024
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 599,543
FILING DATE: 18-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 579,865
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 569,920
FILING DATE: 20-AUG-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 483,913
FILING DATE: 22-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 422,613
FILING DATE: 17-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 315,342
FILING DATE: 23-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 232,630
FILING DATE: 15-AUG-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 179,460
FILING DATE: 08-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER, EDMUND R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-001CP6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 1723 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: HIPPOCAMPUS
FEATURE:
NAME/KEY: CDS
LOCATION: 490..1696
OTHER INFORMATION: /function="OSTEOGENIC PROTEIN"
OTHER INFORMATION: /product="hop2-p"
OTHER INFORMATION: /note="hop2 (cDNA)"
US-08-147-023-28

Query Match      10.8%; Score 28; DB 1; Length 1723;
Best Local Similarity 47.0%; Pred. No. 2.2;
Matches 47; Conservative 8; Mismatches 45; Indels 0; Gaps 0;

QY 1 aguuuugcgacgaugugaucaagaacgguuucgguguaagccauugucugaagggaug 60
    ||:| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 808 AGCTTCGTTAAACATGATGTGAGAGAGACCGCTGCGCCCTGGGCCACACGAGAGCCCATTTGGAAG 867

QY 61 acguccuucucugucuaucacaaacccgucucggugug 100
    | | : | | | | | | | | | | | | | | | | | | |
DB 866 GAGTTCGCTTGACCTGACCCAGATCCCGGCTGGGAGG 907

RESULT 7
US-08-206-864-3
; Sequence 3, Application US/08206864
```

```
; Patent No. 5610021
; GENERAL INFORMATION:
; APPLICANT: RUEGER, DAVID C
; APPLICANT: JONES, WILLIAM K
; APPLICANT: TUCKER, RONALD F
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; RECOMBINANT OSTEOGENIC PROTEIN PRODUCTION
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR/CREATIVE BIOMOLECULES,
; INC.
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/206,864
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/027,070
; FILING DATE: 04-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/841,646
; FILING DATE: 21-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-096
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/7000
; TELEFAX: 617/248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: HIPPOCAMPI
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 490..1696
; OTHER INFORMATION: /function="OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product="hop2-pp"
; OTHER INFORMATION: /note="hop2 (CDNA)"
US-08-206-864-3
```

```
Query Match 10.8%; Score 28; DB 1; Length 1723;
Best Local Similarity 47.0%; Pred. No. 2.2;
Matches 47; Conservative 8; Mismatches 45; Indels 0; Gaps 0;

QY 1 aguuugcgaccgauggaucagaaacgguuugggugaaagcgaugucugaagggaug 60
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 808 AGCTTCGTTACATGATGAGAGACCGCTGCGGCCACAGAGACCCCATTTGAAG 867

QY 61 acguccuucugucucacacacaaacgguucggugug 100
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 868 GAGTTCGCTTGACCTGACCCAGATCCCGGCGTGGGAGG 907
```

```
RESULT 8
US-08-278-729A-20
; Sequence 20, Application US/08278729A
; Patent No. 5650276
; GENERAL INFORMATION:
; APPLICANT: SMART, JOHN
; APPLICANT: OPPERMAN, HERMAN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: MORPHOGENIC PROTEIN SCREENING METHOD
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,729A
; FILING DATE: 20-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-058CPFW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-9001
; TELEFAX: (508) 435-6951
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 490..1695
US-08-278-729A-20
```

```
Query Match 10.8%; Score 28; DB 1; Length 1723;
Best Local Similarity 47.0%; Pred. NO. 2.2;
Matches 47; Conservative 8; Mismatches 45; Indels 0; Gaps 0;

QY 1 aguuugcgaccgauggaucagaaacgguuugggugaaagcgaugucugaagggaug 60
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 808 AGCTTCGTTACATGATGAGAGACCGCTGCGGCCACAGAGACCCCATTTGAAG 867

QY 61 acguccuucugucucacacacaaacgguucggugug 100
   ||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 868 GAGTTCGCTTGACCTGACCCAGATCCCGGCGTGGGAGG 907

RESULT 9
US-08-480-528A-7
; Sequence 7, Application US/08480528A
; Patent No. 5652118
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OZKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
```

```

; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,528A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON ESQ., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-076FW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7560
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 490..1696
; OTHER INFORMATION: /function="OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product="hOP2-PP"
; OTHER INFORMATION: /note="hOP2 cDNA"
;
; US-08-480-528A-7
;
; Query Match 10.8%; Score 28; DB 1; Length 1723;
; Best Local Similarity 47.0%; Pred. No. 2.2;
; Matches 47; Conservative 8; Mismatches 45; Indels 0; Gaps 0;
;
; QY 1 aguuuggcacccaugugaucaaacgcuuugcguguaagccaugucugaaagggaug 60
; ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; DB 808 ACCTTCGTTAACTGATGTTGAGCGAGACCGTCCCTGGCCACCAGGAGCCCATTTGGAAG 867
;
; QY 61 acguccuucugugcucauacacaacacgucucugugug 100
; ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; DB 868 GAGTTCGCGCTTGACCTGACCCAGATCCCGCGCTGGGGAGG 907
;
; RESULT 10
; US-08-479-666-7
; Sequence 7, Application US/08479666
; Patent No. 5652337
; GENERAL INFORMATION:
; APPLICANT: OPPERMAN, HERMANN
; APPLICANT: OKAYNAK, ENGIN
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: PANG, ROY H.L.
; APPLICANT: COHEN, CHARLES M.
; TITLE OF INVENTION: OP3-INDUCED MORPHOGENESIS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA

```

```

; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,666
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FENTON ESQ., GILLIAN M.
; REGISTRATION NUMBER: 36,508
; REFERENCE/DOCKET NUMBER: CRP-076DV
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 435-6951
; TELEFAX: (508) 435-9001
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1723 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 490..1696
; OTHER INFORMATION: /function="OSTEOGENIC PROTEIN"
; OTHER INFORMATION: /product="hOP2-PP"
; OTHER INFORMATION: /note="hOP2 cDNA"
;
; US-08-479-666-7
;
; Query Match 10.8%; Score 28; DB 1; Length 1723;
; Best Local Similarity 47.0%; Pred. No. 2.2;
; Matches 47; Conservative 8; Mismatches 45; Indels 0; Gaps 0;
;
; QY 1 aguuuggcacccaugugaucaaacgcuuugcguguaagccaugucugaaagggaug 60
; ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; DB 808 ACCTTCGTTAACTGATGTTGAGCGAGACCGTCCCTGGCCACCAGGAGCCCATTTGGAAG 867
;
; QY 61 acguccuucugugcucauacacaacacgucucugugug 100
; ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
; DB 868 GAGTTCGCGCTTGACCTGACCCAGATCCCGCGCTGGGGAGG 907
;
; RESULT 11
; US-08-155-343A-20
; Sequence 20, Application US/08155343A
; Patent No. 5656593
; GENERAL INFORMATION:
; APPLICANT: KUBERASAMPATH, THANGAVEL
; APPLICANT: RUEGER, DAVID C.
; APPLICANT: OPPERMAN, HERMAN
; APPLICANT: COHEN, CHARLES M.
; APPLICANT: PANG, ROY H.L.
; TITLE OF INVENTION: MORPHOGENIC-INDUCED PERIODONTAL TISSUE
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES
; STREET: 45 SOUTH STREET
; CITY: HOPKINTON
; STATE: MA
; COUNTRY: USA
; ZIP: 01748
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

```

ATTORNEY/AGENT INFORMATION:

```

; TYPE: nucleic acid

```









GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 15:22:07 ; Search time 2783.16 Seconds  
(without alignments)  
1631.001 Million cell updates/sec

Title: US-09-587-653-1

Perfect score: 259  
1 aqunugcgaccacgugugga.....cccgcuuggaauaaaacu 259

Sequence:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 17159718 seqs, 8763200856 residues

Total number of hits satisfying chosen parameters: 34319436

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pending\_Patents\_NA\_Main:\*

1:	/cgn2_6/p/oddata/2/pna/US07.COMB.seq:*
2:	/cgn2_6/p/oddata/2/pna/US06.COMB.seq:*
3:	/cgn2_6/p/oddata/2/pna/US07.COMB.seq:*
4:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
5:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
6:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
7:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
8:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
9:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
10:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
11:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
12:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
13:	/cgn2_6/p/oddata/2/pna/US08.COMB.seq:*
14:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
15:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
16:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
17:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
18:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
19:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
20:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
21:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
22:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
23:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
24:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
25:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
26:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
27:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
28:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
29:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
30:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
31:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
32:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
33:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
34:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
35:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
36:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
37:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
38:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
39:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
40:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
41:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
42:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*
43:	/cgn2_6/p/oddata/2/pna/US09.COMB.seq:*

Result No.	Score	Query Match	Length	ID	Description
44:	259	100.0	259	US-09-587-653-1	Sequence 1, App11
45:	259	100.0	9399	US-09-587-653-2	Sequence 2, App11
46:	259	100.0	9399	US-09-742-653-1	Sequence 1, App11
47:	259	100.0	9399	US-09-137-694-1	Sequence 1, App11
48:	257.4	99.4	259	US-09-579-302-2	Sequence 3, App11
49:	257.4	99.4	259	US-09-579-302-3	Sequence 3, App11
50:	257.4	99.4	259	US-09-579-302-4	Sequence 3, App11
51:	257.4	99.4	259	US-09-579-302-5	Sequence 3, App11
52:	257.4	99.4	259	US-09-579-302-6	Sequence 3, App11
53:	257.4	99.4	259	US-09-579-302-7	Sequence 3, App11
54:	257.4	99.4	259	US-09-579-302-8	Sequence 3, App11
55:	257.4	99.4	259	US-09-579-302-9	Sequence 3, App11
56:	257.4	99.4	259	US-09-579-302-10	Sequence 3, App11
57:	257.4	99.4	259	US-09-579-302-11	Sequence 3, App11
58:	257.4	99.4	259	US-09-579-302-12	Sequence 3, App11
59:	257.4	99.4	259	US-09-579-302-13	Sequence 3, App11
60:	257.4	99.4	259	US-09-579-302-14	Sequence 3, App11
61:	257.4	99.4	259	US-09-579-302-15	Sequence 3, App11
62:	257.4	99.4	259	US-09-579-302-16	Sequence 3, App11
63:	257.4	99.4	259	US-09-579-302-17	Sequence 3, App11
64:	257.4	99.4	259	US-09-579-302-18	Sequence 3, App11
65:	257.4	99.4	259	US-09-579-302-19	Sequence 3, App11

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	259	100.0	259	US-09-587-653-1	Sequence 1, App11
2	259	100.0	9399	US-09-587-653-2	Sequence 2, App11
3	259	100.0	9399	US-09-742-653-1	Sequence 1, App11
4	259	100.0	9399	US-09-137-694-1	Sequence 1, App11
5	257.4	99.4	259	US-09-579-302-2	Sequence 3, App11
6	257.4	99.4	259	US-09-579-302-3	Sequence 3, App11
7	257.4	99.4	259	US-09-579-302-4	Sequence 3, App11
8	257.4	99.4	259	US-09-579-302-5	Sequence 3, App11
9	257.4	99.4	259	US-09-579-302-6	Sequence 3, App11
10	257.4	99.4	259	US-09-579-302-7	Sequence 3, App11
11	257.4	99.4	259	US-09-579-302-8	Sequence 3, App11
12	257.4	99.4	259	US-09-579-302-9	Sequence 3, App11
13	257.4	99.4	259	US-09-579-302-10	Sequence 3, App11
14	257.4	99.4	259	US-09-579-302-11	Sequence 3, App11
15	257.4	99.4	259	US-09-579-302-12	Sequence 3, App11
16	257.4	99.4	259	US-09-579-302-13	Sequence 3, App11
17	257.4	99.4	259	US-09-579-302-14	Sequence 3, App11
18	257.4	99.4	259	US-09-579-302-15	Sequence 3, App11
19	257.4	99.4	259	US-09-579-302-16	Sequence 3, App11
20	257.4	99.4	259	US-09-579-302-17	Sequence 3, App11
21	257.4	99.4	259	US-09-579-302-18	Sequence 3, App11
22	257.4	99.4	259	US-09-579-302-19	Sequence 3, App11
23	257.4	99.4	259	US-09-579-302-20	Sequence 3, App11
24	257.4	99.4	259	US-09-579-302-21	Sequence 3, App11
25	257.4	99.4	259	US-09-579-302-22	Sequence 3, App11
26	257.4	99.4	259	US-09-579-302-23	Sequence 3, App11
27	257.4	99.4	259	US-09-579-302-24	Sequence 3, App11
28	257.4	99.4	259	US-09-579-302-25	Sequence 3, App11
29	257.4	99.4	259	US-09-579-302-26	Sequence 3, App11
30	257.4	99.4	259	US-09-579-302-27	Sequence 3, App11
31	257.4	99.4	259	US-09-579-302-28	Sequence 3, App11
32	257.4	99.4	259	US-09-579-302-29	Sequence 3, App11
33	257.4	99.4	259	US-09-579-302-30	Sequence 3, App11
34	257.4	99.4	259	US-09-579-302-31	Sequence 3, App11
35	257.4	99.4	259	US-09-579-302-32	Sequence 3, App11
36	257.4	99.4	259	US-09-579-302-33	Sequence 3, App11
37	257.4	99.4	259	US-09-579-302-34	Sequence 3, App11
38	257.4	99.4	259	US-09-579-302-35	Sequence 3, App11
39	257.4	99.4	259	US-09-579-302-36	Sequence 3, App11
40	257.4	99.4	259	US-09-579-302-37	Sequence 3, App11
41	257.4	99.4	259	US-09-579-302-38	Sequence 3, App11
42	257.4	99.4	259	US-09-579-302-39	Sequence 3, App11
43	257.4	99.4	259	US-09-579-302-40	Sequence 3, App11





```
RESULT 7
US-09-579-302-4
; Sequence 4, Application US/09579302
; GENERAL INFORMATION:
; APPLICANT: Traboni, Cinzia
; TITLE OF INVENTION: Novel GBV sequence
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/579,302
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: GB 9912432.3
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 259
; TYPE: DNA
; ORGANISM: GBV-B-like virus
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1)..(259)
; OTHER INFORMATION: cDNA of positive strand 3'UTR of GBV-B
US-09-579-302-4
```

```
Query Match          99.4%; Score 257.4; DB 22; Length 259;
Best Local Similarity 76.1%; Pred. No. 1.2e-73;
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 aguuugcgaccgaugugaucaacccguuucggugagaagccaugugucgaagggaug 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 agtttggcagcaatggtgatacgaacccgttcgggaggaagcaatgctgaagggaatg 60

QY 61 acgucccuuugcucuaucacaaacccgucucggugugugaggaagucugucugug 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 acgtccctcttgctatccacaaacgctcgggtggtgaggaagtccttgctgtgt 120

QY 121 gggaagcagucaguanauuuccgucugugugagcgcucacgaaguanuugccgc 180
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 gggaagcagctagataatcccgctgtgtgtgtagcactacgaagctactgtcgcg 180

QY 181 ugugcagagcaguanuuccaaggcugaccccgguuuuuguccaagcgaggaagcaacc 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 tgtgcagagcgtagtaaccagagcgtgaccccggttcttctccaaagcgaggaagcaacc 240

QY 241 ccgcugugaaunaaaaacu 259
   |||||:|||||:|||||:
Db 241 ccgcttggaattaaaaact 259
```

```
RESULT 8
US-09-579-302-5/c
; Sequence 5, Application US/09579302
; GENERAL INFORMATION:
; APPLICANT: Traboni, Cinzia
; TITLE OF INVENTION: Novel GBV sequence
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/579,302
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: GB 9912432.3
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 259
; TYPE: DNA
; ORGANISM: GBV-B-like virus
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1)..(259)
; OTHER INFORMATION: cDNA of negative strand 3'UTR of GBV-B
US-09-579-302-5
```

```
Query Match          99.4%; Score 257.4; DB 22; Length 259;
Best Local Similarity 76.1%; Pred. No. 1.2e-73;
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;

QY 1 aguuugcgaccgaugugaucaacccguuucggugagaagccaugugucgaagggaug 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 agtttggcagcaatggtgatacgaacccgttcgggaggaagcaatgctgaagggaatg 200

QY 61 acgucccuuugcucuaucacaaacccgucucggugugugaggaagucugucugug 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 acgtccctcttgctatccacaaacgctcgggtggtgaggaagtccttgctgtgt 140

QY 121 gggaagcagucaguanauuuccgucugugugagcgcucacgaaguanuugccgc 180
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 gggaagcagctagataatcccgctgtgtgtgtagcactacgaagctactgtcgcg 80

QY 181 ugugcagagcaguanuuccaaggcugaccccgguuuuuguccaagcgaggaagcaacc 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 79 tgtgcagagcgtagtaaccagagcgtgaccccggttcttctccaaagcgaggaagcaacc 20

QY 241 ccgcugugaaunaaaaacu 259
   |||||:|||||:|||||:
Db 19 ccgcttggaattaaaaact 1
```

```
RESULT 9
US-09-579-302-1
; Sequence 1, Application US/09579302
; GENERAL INFORMATION:
; APPLICANT: Traboni, Cinzia
; TITLE OF INVENTION: Novel GBV sequence
; FILE REFERENCE: Mewburn
; CURRENT APPLICATION NUMBER: US/09/579,302
; CURRENT FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: GB 9912432.3
; PRIOR FILING DATE: 1999-05-27
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 309
; TYPE: DNA
; ORGANISM: GBV-B-like virus
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1)..(309)
; OTHER INFORMATION: cDNA complementary to the last 49 published 3'UTR
; OTHER INFORMATION: nucleotides of GBV-B plus a novel nucleotide (C)
; OTHER INFORMATION: in position 44 and 259 novel nucleotide sequence
; OTHER INFORMATION: at the 3' end.
US-09-579-302-1
```

```
Query Match          99.4%; Score 257.4; DB 22; Length 309;
Best Local Similarity 76.1%; Pred. No. 1.3e-73;
Matches 197; Conservative 61; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 aguuugcgaccgaugugaucaacccguuucggugagaagccaugugucgaagggaug 60
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 51 agtttggcagcaatggtgatacgaacccgttcgggaggaagcaatgctgaagggaatg 110

QY 61 acgucccuuugcucuaucacaaacccgucucggugugugaggaagucugucugug 120
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 111 acgtccctcttgctatccacaaacgctcgggtggtgaggaagtccttgctgtgt 170

QY 121 gggaagcagucaguanauuuccgucugugugagcgcucacgaaguanuugccgc 180
   |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 171 gggaagcagctagataatcccgctgtgtgtgtagcactacgaagctactgtcgcg 230

QY 181 ugugcagagcaguanuuccaaggcugaccccgguuuuuguccaagcgaggaagcaacc 240
   ||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 231 tgtgcagagcgtagtaaccagagcgtgaccccggttcttctccaaagcgaggaagcaacc 290

QY 241 ccgcugugaaunaaaaacu 259
```

FOR INTERF

[illegible]

Query Match	13.33;	Score 34.4;	DB 29;	Length 790;
Best Local Similarity	45.14;	Pred. No. 2.3;		
Matches 64;	Conservative 10;	Mismatches 68;	Indels 0;	Gaps

[illegible]

```

; ORGANISM: Homo sapiens
US-09-306-350A-23940

Query Match      12.4%  Score 33.2;  DB 17;  Length 429;
Best Local Similarity 44.4%;  Pred No. 10;
Matches 52;  Conservative 12;  Mismatches 55;  Indels 0;  Gaps 0;

QY      6  ggcgcacgaugugagucagaaccgguucgggugaagccaugucugaaggggaugacguc 65
      |||  :|:|  |  |||||  |  :  :|||  |||  |||  |||  |
Db      116  gacgctgaatgctcttgagcagaacagactgcgcatggaagaaatctggaagaagaatctt 175
      |||  :|:|  |  |||||  |  :  :|||  |||  |||  |||  |

QY      66  ccuucugugucucacacaaaacgcugucugggugugagagcugucugcgucugugug 122
      |||  |||  |  ||||  |  :  |||  |  :  |||  |  |||
Db      176  tcgaagagctcctaaagaataatcctctcttgaggcaagagctcttgagctgctcg 232

```

```

Query Match          12.6%; Score 32.6; DB 17; Length 413;
Best Local Similarity 32.5%; Pred. No. 7.5;
Matches   62; Conservative    30; Mismatches   99; Indels     0; Gaps      0;

QY       3 uuuugcgaccacgungugagucagaacguuuccggugugaagcccaugugcuaaggugggaugac 62
         ::|||:::||||:::||::|||::|||::|||::|||
Db        171 ttgtgcacatgcccataatatattttcttactttgttagcacatatctfgaaagaataac 230
               |||::|||::|||::|||::|||::|||::|||::|||

QY       63 guccccuucgugcucaucaacaacaaaacgcugcgguguguggaaggauccugcugugug 122
         :::|||::|||::|||::|||::|||::|||::|||::|||
Db        221 atttttaagccttcaccacacataagacatatcatgtctaagatlgtytgtytgtgtgt 290
               |::|||::|||::|||::|||::|||::|||::|||

QY       123 gaagcagucaguanaauucccgugugugugugagcgccucacgcagacuauuugcgcug 182
         |::|||::|||::|||::|||::|||::|||::|||::|||
Db        291 ggtatgtatgtatagtgcggagagagatagtgtgttggaacagacataaggaagcgg 350
               |::|||::|||::|||::|||::|||::|||::|||

QY       183 ugacagagcgua 193
         |::|||::|||::|||::|||::|||::|||::|||

Db        351 ggaggaactgga 361
               |::|||::|||::|||::|||::|||::|||::|||

RESULT   15
US-09-306-350A-23940
: Sequence 23940, Application US/09306350A
: GENERAL INFORMATION:
: APPLICANT: Dickson, Mark C.
: APPLICANT: Drmanac, Radoje T.
: APPLICANT: Jones, Lee W.
: APPLICANT: Labat, Ivan
: APPLICANT: Stache-Crain, Birgit
: TITLE OF INVENTION: Novel Nucleic Acid Sequences Obtained
: TITLE OF INVENTION: From Various cDNA Libraries
: FILE REFERENCE: 20411-776
: CURRENT APPLICATION NUMBER: US/09/306,350A
: CURRENT FILING DATE: 1999-05-07
: NUMBER OF SEQ ID NOS: 41304
: SOFTWARE: PL CT_1 Version 1.1
: SEO ID NO 23940
: LENGTH: 429
: TYPE: DNA
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 15:47:52 ; Search time 438.55 Seconds  
(without alignments)  
1487.656 Million cell updates/sec

Title: US-09-587-653-1

Perfect score: 259

Sequence: 1 aguuugcgaccuagugaga.....cccgcuuggaauuaaaca 259

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2186606 seqs, 1259481600 residues

Total number of hits satisfying chosen parameters: 4373212

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending\_Patents\_NA\_New:\*

1: /cgn2\_6/ptodata/2/pna/PCT\_NEW\_COMB.seq:\*

2: /cgn2\_6/ptodata/2/pna/US06\_NEW\_COMB.seq:\*

3: /cgn2\_6/ptodata/2/pna/US08\_NEW\_COMB.seq:\*

4: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*

5: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*

6: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*

7: /cgn2\_6/ptodata/2/pna/US09\_NEW\_COMB.seq:\*

8: /cgn2\_6/ptodata/2/pna/US10\_NEW\_COMB.seq:\*

9: /cgn2\_6/ptodata/2/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32.6	12.6	16943	US-10-015-127-485	Sequence 465, App
2	31	12.0	1095	US-10-000-897-36	Sequence 36, App
3	30	11.6	381	US-09-849-526A-10014	Sequence 10014, A
4	30	11.6	390	US-09-975-673A-5933	Sequence 5933, App
5	30	11.6	49999	US-09-647-377-16	Sequence 16, App
6	29.8	11.5	503	US-09-873-402A-78798	Sequence 78798, A
7	29.8	11.5	533	US-09-873-402A-9421	Sequence 9421, App
8	29.8	11.5	702	US-09-865-439A-74213	Sequence 74213, A
9	29.4	11.4	385	US-09-669-817A-38740	Sequence 38740, A
10	29.4	11.4	460	US-10-029-386-25170	Sequence 25170, A
11	29.4	11.4	642	US-09-865-439A-79096	Sequence 79096, A
12	29.2	11.3	148	US-09-963-299-95	Sequence 95, App
13	29.2	11.3	31782	US-09-815-264-71589	Sequence 71589, A
14	29	11.2	36991	US-09-815-264-71589	Sequence 71589, A
15	29	11.2	166649	US-10-021-698-704	Sequence 704, App
16	29	11.2	169752	US-10-021-698-703	Sequence 703, App
17	28.8	11.1	348	US-09-669-817A-25063	Sequence 25063, A
18	28.8	11.1	549	US-10-021-698-3838	Sequence 3838, App
19	28.8	11.1	891	US-09-708-427-18324	Sequence 18324, A
20	28.6	11.0	393	US-09-925-564-6641	Sequence 6641, App
21	28.6	11.0	452	US-09-865-419A-22394	Sequence 22394, A
22	28.6	11.0	1236	US-08-708-427-2952	Sequence 2952, App
23	28.4	11.0	3384	US-08-341-261-1331	Sequence 1331, App
24	28.4	11.0	5384	US-60-341-261-331	Sequence 331, App

25	28.2	10.9	347	US-09-927-875A-978	Sequence 978, App
26	28.2	10.9	1501	US-09-815-264-19757	Sequence 19757, A
27	28.2	10.9	5493	US-09-760-446A-2499	Sequence 2499, App
28	28.2	10.9	5717	US-09-760-446A-2500	Sequence 2500, App
29	28.2	10.9	32372	US-09-815-264-58867	Sequence 58867, A
30	28	10.8	362	US-09-669-817A-22274	Sequence 22274, A
31	28	10.8	414	US-09-675-84A-2905	Sequence 2905, App
32	28	10.8	1723	US-09-148-925A-28	Sequence 28, App
33	28	10.8	15650	US-09-927-799A-2074	Sequence 2074, App
34	28	10.8	5650	US-09-815-264-59843	Sequence 59843, A
35	28	10.8	16440	US-09-815-264-59843	Sequence 59843, A
36	28	10.8	36711	US-09-815-264-59843	Sequence 63142, A
37	28	10.8	169998	US-09-933-524A-86892	Sequence 86892, A
38	27.8	10.7	229	US-09-985-678-77775	Sequence 77775, A
39	27.8	10.7	389	US-09-933-524A-86892	Sequence 86892, A
40	27.8	10.7	400	US-09-985-678-98738	Sequence 98738, A
41	27.8	10.7	588	US-09-865-439A-81324	Sequence 81324, A
42	27.8	10.7	609	US-09-873-402A-84520	Sequence 84520, A
43	27.8	10.7	609	US-10-034-934-43	Sequence 43, App
44	27.8	10.7	839	US-09-865-439A-20253	Sequence 20253, A
45	27.8	10.7	1498	US-09-898-888A-14368	Sequence 14368, A

## ALIGNMENTS

```

RESULT 1
US-10-015-127-485/c
; Sequence 485, Application US/10015127
; GENERAL INFORMATION:
; APPLICANT: Bower, Stanley G.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: Sphingomonas elodea genome sequences and uses thereof
; FILE REFERENCE: 38-10713806B
; CURRENT FILING DATE: 2001-10-127
; PRIOR APPLICATION NUMBER: US/10-015-127
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 14357
; SEQ ID NO 485
; LENGTH: 16943
; TYPE: DNA
; ORGANISM: Sphingomonas elodea
US-10-015-127-485

Query Match 12.6%; Score 32.6; DB 8; Length 16943;
Best Local Similarity 50.6%; Pred. No. 2.8;
Matches 40; Conservative 10; Mismatches 29; Indels 0; Gaps 0;

Oy 56 ggaugagcuccuucgucacacaaacccgucgugugagagucucgc 115
    |||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 9744 GCGAGCGCTCCGGTCCAGTCGCGCAGATCAGCCGCGATGATGATGAGGATCGCGC 9685

Oy 116 uugugugagagcagucagu 134
    |:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 9684 TGCATGCGAGCCACTGACT 9666

RESULT 2
US-10-000-897-36
; Sequence 36, Application US/10000897
; GENERAL INFORMATION:
; APPLICANT: Schueler, P.
; TITLE OF INVENTION: METHODS AND REAGENTS FOR IDENTIFYING RARE FETAL CELLS IN THE
; FILE REFERENCE: 11012-004-999
; CURRENT APPLICATION NUMBER: US/10-000-897
; PRIOR FILING DATE: 2001-11-15
; PRIOR FILING DATE: 2000-11-15
; NUMBER OF SEQ ID NOS: 78

```



```
; APPLICANT: Varagona, Marguerite J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51934)B
; CURRENT APPLICATION NUMBER: US/09/873,402A
; CURRENT FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,830
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 90966
; SEQ ID NO 78798
; LENGTH: 503
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3632-059-Q6-N6-C1
US-09-873-402A-78798
```

```
Query Match          11.5%; Score 29.8; DB 5; Length 503;
Best Local Similarity 47.7%; Pred. No. 8.7;
Matches 31; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
```

```
QY 30 uuucggugagaagccauugucuaaggggaugacgucuccuucugucuaucacacaaacc 89
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 67 TTTCGTGTGTACCATGAGCTGCTGCATTCGCTCCCTCTGTATCCAAACAACAACATT 8
```

```
QY 90 guucuc 94
      |::|:|
Db 7 GTCCTC 3
```

## RESULT 7

```
; US-09-873-402A-9421/C
; Sequence 9421, Application US/09873402A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
; APPLICANT: Hardeman, Kristine J.
; APPLICANT: Varagona, Marguerite J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51934)B
; CURRENT APPLICATION NUMBER: US/09/873,402A
; CURRENT FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/209,830
; PRIOR FILING DATE: 2000-06-06
; NUMBER OF SEQ ID NOS: 90966
; SEQ ID NO 9421
; LENGTH: 533
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3591-104-Q1-N6-A9
US-09-873-402A-9421
```

```
Query Match          11.5%; Score 29.8; DB 5; Length 533;
Best Local Similarity 47.7%; Pred. No. 8.8;
Matches 31; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
```

```
QY 30 uuucggugagaagccauugucuaaggggaugacgucuccuucugucuaucacacaaacc 89
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 101 TTTCGTGTGTACCATGAGCTGCTGCATTCGCTCCCTCTGTATCCAAACAACAACATT 42
```

```
QY 90 guucuc 94
      |::|:|
Db 41 GTCCTC 37
```

## RESULT 8

```
; US-09-865-439A-74213/C
; Sequence 74213, Application US/09865439A
; GENERAL INFORMATION:
; APPLICANT: Edgerton, Michael D
```

```
; APPLICANT: Hardeman, Kristine J.
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51936)B
; CURRENT APPLICATION NUMBER: US/09/865,439A
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/207,458
; PRIOR FILING DATE: 2000-05-30
; NUMBER OF SEQ ID NOS: 119126
; SEQ ID NO 74213
; LENGTH: 702
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: jC-zmsHLIB3587P031f02a1
US-09-865-439A-74213
```

```
Query Match          11.5%; Score 29.8; DB 7; Length 702;
Best Local Similarity 47.7%; Pred. No. 9.5;
Matches 31; Conservative 12; Mismatches 22; Indels 0; Gaps 0;
```

```
QY 30 uuucggugagaagccauugucuaaggggaugacgucuccuucugucuaucacacaaacc 89
      ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 141 TTTCGTGTGTACCATGAGCTGCTGCATTCGCTCCCTCTGTATCCAAACAACAACATT 82
```

```
QY 90 guucuc 94
      |::|:|
Db 81 GTCCTC 77
```

## RESULT 9

```
; US-09-669-817A-38740
; Sequence 38740, Application US/09669817A
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: Ruan, Yijun G.
; APPLICANT: Wallick, C. Kevin
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-21(51469)B
; CURRENT APPLICATION NUMBER: US/09/669,817A
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/156,951
; PRIOR FILING DATE: 1999-09-30
; NUMBER OF SEQ ID NOS: 43701
; SEQ ID NO 38740
; LENGTH: 385
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: uc-osrom20205h07b1
US-09-669-817A-38740
```

```
Query Match          11.4%; Score 29.4; DB 7; Length 385;
Best Local Similarity 48.5%; Pred. No. 11;
Matches 50; Conservative 7; Mismatches 46; Indels 0; Gaps 0;
```

```
QY 2 uuucggagccauugucuaagccguuuucgguguaagccauugucuaaggggauga 61
      |::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 133 ggttggtggtccatggcggagctgaagcggctgtctcgagagagcgcgacctgaagcgatcg 192
```

```
QY 62 cuuccuucugucuaucacacaaaccguucucgguguguga 104
      |::|:|::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db 193 agcgcatcggggcgccactccacatccggggggtcgggctgga 235
```

## RESULT 10

```
; US-10-029-386-25170
; Sequence 25170, Application US/10029386
; GENERAL INFORMATION:
```



```

; TYPE: DNA
; ORGANISM: Human

```

```
; NAME/KEY: misc_feature
;; LOCATION: (1)..(166649)
;; OTHER INFORMATION: n = A,T,C or G
```

Query Match

9	gacgaugugugaucagaaacgguuucggguguaagcgaugucucuaaagggaugacgucucu	68
QY	gacgaugugugaucagaaacgguuucggguguaagcgaugucucuaaagggaugacgucucu	68
Matches	44; Conservative 12; Mismatches 45; Indels 0; Gaps 0;	

MD 984/8 GGGCCCTGAGGAGTCTTCCGGAGTCTTAAAAACCAAGGTCACAGCGCGGCTCCATGACCG 984/12  
 QY 69 ucuugcucacuccaacaataaccgucucgugugugugugagggaguu 109

DD 50410 AC1011CA1100CGAAGCCACACGCG1GAG1CAGGAG1 503/6

Job time: 5062 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 16, 2002, 14:00:57 ; Search time 1527.54 Seconds  
(without alignments)  
1821.986 Million cell updates/sec

Title: US-09-587-653-1

Perfect score: 259  
Sequence: 1 aguuugcgaccgaugunga.....cccgcuuggaauaanaacu 259

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estln:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estlda:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hlc:\*  
10: gb\_estl:\*  
11: gb\_estc2:\*  
12: gb\_hlc:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.8	13.4	670	13	AZ962384 2M0231L1L1
2	33.8	13.1	1092	13	AL347370 Tetradon
3	33.2	12.6	1508	10	BE615183
4	33	12.7	1054	11	BG687472
5	32.6	12.6	287	10	AV072177
6	32.6	12.6	561	13	AO657755 Sheared D
7	32	12.4	283	10	BB385246
8	31.6	12.2	1022	10	AL568674
9	31.4	12.1	321	10	BB500470
10	31.2	12.0	251	10	BB425694
11	31.2	12.0	281	10	BB421037
12	31.2	12.0	712	11	BF577610

13	31	12.0	275	10	BE647687
14	31	12.0	548	10	AM361210
15	31	12.0	586	13	BH110789
16	31	12.0	918	13	CNS02JF4
17	31	12.0	1907	12	AK005169
18	31	12.0	469	11	BG639798
19	30.6	11.8	533	10	A1533127
20	30.6	11.8	555	10	A1533553
21	30.6	11.8	654	10	AA951616
22	30.6	11.8	730	10	AA816991
23	30.6	11.8	743	10	BE407858
24	30.6	11.8	749	13	CNS04ICM
25	30.6	11.8	1410	11	BG167510
26	30.4	11.7	259	13	CNS04GUT
27	30.4	11.7	278	10	BB365538
28	30.4	11.7	421	10	AU181868
29	30.4	11.7	468	11	BG730062
30	30.4	11.7	716	11	BF615851
31	30.4	11.7	750	11	BF613405
32	30.4	11.7	861	11	BI217114
33	30.4	11.7	1135	11	BG36385
34	30.2	11.7	279	10	BB382397
35	30.2	11.7	499	11	BG164682
36	30.2	11.7	1072	13	CNS05C84
37	30	11.6	262	10	BB416099
38	30	11.6	314	10	BB162728
39	30	11.6	529	11	BG515211
40	30	11.6	546	11	BG512901
41	30	11.6	560	11	BG633101
42	30	11.6	566	11	BG815751
43	30	11.6	575	11	BG512819
44	30	11.6	655	10	AV703583
45	30	11.6	683	10	AV705255

#### ALIGNMENTS

RESULT 1  
AZ962384 LOCUS 670 bp DNA GSS 27-APR-2001  
DEFINITION 2M0231L1L1F Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
clone UUGC2M0231L1L1 F, DNA sequence.  
ACCESSION AZ962384  
VERSION 1 GI:13833611  
KEYWORDS GSS.  
ORGANISM house mouse.  
SOURCE Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.  
REFERENCE 1 (bases 1 to 670)  
AUTHORS Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamli,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,  
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.  
and Wright,D., Weiss,R.  
TITLE Mouse whole genome scaffolding with paired end reads from 10kb  
plasmid inserts  
JOURNAL Unpublished (2000)  
COMMENT Contact: Robert B. Weiss  
University of Utah  
Genome Center  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT  
84112, USA  
Tel.: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0231 row: L column: 11  
Seq primer: CGTGTAAACGACGCCACG  
Class: plasmid ends  
High quality sequence stop: 670.  
Location/Qualifiers 1..670

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U062M023111"
/clone_lib="Mouse 10kb plasmid U062M library"
/sex="female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (914732114|9b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      168 a      151 c      180 g      171 t
ORIGIN

```

```

Query Match      13.4%; Score 34.8; DB 13; Length 670;
Best Local Similarity 39.7%; Pred. No. 2.7;
Matches 50; Conservative 19; Mismatches 57; Indels 0; Gaps 0;

```

```

QY 133 guaaauuuccgugugugugagcgccuacgaguanuuugcugugagagcgcu 192
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 351 GTGTGACTGTGCACTGTGTGTGTGACCAACCTCTGTCTGTGCTGCCACATGT 410

```

```

QY 193 aguacaaaggcugcagcccgguuuuuccaagcgagggcaaccccgucugaaau 252
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 411 AAAATGAAGATAGAGCCCTCAGGTGTGTACGAATGAATGAACACATGTGGGAAA 470

```

```

QY 253 aaaaac 258
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 471 GATTAAC 476

```

```

RESULT 2
CNS05POX/c 1092 bp DNA GSS 26-MAY-2000
LOCUS Tetracodon nigroviridis genome survey sequence T7 end of clone
045618 of library A from Tetracodon nigroviridis, genomic survey
sequence.
AL347370.1 GI:8241140
VERSION AL347370.1 GI:8241140
KEYWORDS GSS; genome survey sequence.
SOURCE Tetracodon nigroviridis.
ORGANISM Tetracodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorphi; Acanthopterygii; Perciformes; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 1092)
Roest-Crollius,H., Jallion,O., Dasilva,C., Fizames,C., Fisher,C.,
Bonneau,L., Billault,A., Quetler,F., Saurin,W., Bernot,A. and
Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetracodon nigroviridis
Unpublished
2 (bases 1 to 1092)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetler,F.,
Saurin,W. and Weissenbach,J.

```

```

TITLE Human gene number estimate provided by genome wide analysis using
Tetracodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 1092)
AUTHORS
TITLE Direct Submission
COMMENT Submitted (12-Apr-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetracodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetracodon.
FEATURES
source
1..1092
/organism="Tetracodon nigroviridis"
/db_xref="taxon:99883"
/clone="045618"
/clone_lib="A"
/notes="Genoscope sequence ID : COA045BD09C1-end : T7"
BASE COUNT      281 a      240 c      234 g      313 t      24 others
ORIGIN

```

```

Query Match      13.1%; Score 33.8; DB 13; Length 1092;
Best Local Similarity 38.6%; Pred. No. 6.1;
Matches 64; Conservative 19; Mismatches 83; Indels 0; Gaps 0;

```

```

QY 34 gggugagcccaugugucgaggggaugacgucuccuucgucuaacacaaaccguc 93
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 654 GAGTCAGGCTTCTTGTGATGAGAGGTGACACACACGCAACCTCGGAAGGCTTCA 595

```

```

QY 94 cggugugugugagucgucgugugugaggaagcaagcaauaauucccgugugug 153
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 594 TCGTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 535

```

```

QY 154 gtagcgcucacagcagcauuuugcugugagcgagcguaagca 199
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 534 GATCAGACAACTGATTTTATTATTCTTATGCTAATTAATTTATCA 489

```

```

RESULT 3
LOCUS BE615183
DEFINITION BE615183 1508 bp mRNA EST 24-AUG-2000
601281124F1 NIH_MGC_39 Homo sapiens cDNA clone IMAGE:3623261 5',
mRNA sequence.
ACCESSION BE615183
VERSION BE615183.1 GI:9896782
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1508)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L10CM299 row: o column: 06
High quality sequence stop: 307.
Location/Qualifiers
1..1508
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3623261"
/clone_lib="NIH_MGC_39"
/tissue_type="adenocarcinoma"
FEATURES
source

```

```

/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed
by Ling Hong in the laboratory of Gerald M. Rubin
(University of California, Berkeley) using ZAP-cDNA
synthesis kit (Stratagene) and Superscript II RT (Life
Technologies)."
BASE COUNT      316 a      376 c      426 g      388 t      2 others
ORIGIN

```

Query Match	12.8%	Score 33, 2	DB 10	Length 1508
Best Local Similarity	42.7%	Pred. No. 10		
Matches 47, Conservative 15, Mismatches 48			Indels 0	Gaps 0

Qy 88 ccgucucggcgagggagagcucgugcugagaaagcagauaauuccgcug 147  
||| | : | : | | | | | | | : | : | : ||| |  
Db 1084 CCGCCCCAAATTTTGTGGTGC CGCCCGGGGGGGGGGGGATACACCCCGTG 1143

QY 148 ugunugugaacgcucacaagcguauunguccgugucagaagcguaquac 197  
| | | : | | : | : : | : | : | : | :  
Db 1144 GGGGGTTTCTCCTCCGTTGTGTGTCCTGCGGGGGGGTACTAC 1193

RESULT	4	:
BG687472/c		:
LOCUS	BG687472	:
DEFINITION	BG687472 1054 bp mRNA EST	: 01-MAY-2001
	60263944A1 NIH_MGC_59 Homo sapiens CDNA clone IMAGE:4762427 5'	:
	mRNA sequence.	:
*ACCESSION	BG687472	:
VERSION	BG687472.1 GI:13918869	:

ORGANISM	Homo sapiens
NCBI	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homini; Homo.
REFERENCE	1 (bases 1 to 1054)
AUTHORS	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a>
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D.

Plate: LCM1616 row: h column: 12  
High quality sequence stop: 336.

**FEATURES**  
**source**

URES  
source

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4762427"  
 /clone\_lib="NIH\_MGC 59"  
 /tissue\_type="fibroblast/dermoid carcinoma"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /note="Organ: Lung; Vector: pDNR-LIB (Clontech); Site\_1:  
 SfiI (ggcccattgcgcgc); Site\_2: SfiI (ggccattgcgc);  
 Double-stranded cDNA was prepared from cell line RNA.  
 5' and 3' adaptors were used in cloning as follows: 5'  
 adaptor sequence: 5'-CACGGCCATTATGGC-3' and 3'  
 adaptor sequence: 5'-ATTCCTAGGCGGCGCGCATG-dT(30)BN-3'  
 (where B = A, C, or G and N = A, C, G, or T). Average  
 insert size 1.65 kb (range 0.9-4.0 kb). 15/15 colonies  
 contained inserts by PCR. This library was enriched for  
 full-length clones and packaged by Clontech.  
 Laboratories (Palo Alto, CA). Note: this is a NIH\_MGC  
 library."

BASE COUNT	279 a	316 c	157 g	302 f
ORIGIN				

Query Match	12.7%	Score 33;	DB 11;	length 1054;
Best Local Similarity	37.3%	Pred NO. 11;		
Matches 60;	Conservative 21;	Mismatches 80;	Indels 0;	Gaps 0

Qy 90 gucucggugugugagagaguccugcgugugugggaagcagucaguauaaucccugcug 145  
| : | | : | | : | | | | | : | : | :  
Db 810 GTATCGCAAGTTTGAGAAGAGGTGTGTCGCCCATGTGAATTAGCAGGACGAGTAGTGCCTG 751

Oy 150 ugugugacgcgccuacagacgnauuuuugccgcugucagagcguaaguccaaagggcugcac 209  
 : ||| | | | | : | | | | | | | : |  
Db 750 TTTGGTGTGGGCGCCGACAGACCCTTGAGCGTAATGTGTAATAAAAGGATTCG 691

Qy 210 cccgguuuuuuguccaagcagcgagggcaacccccgcuuugaa 250  
||::: ||| ||| : |||  
Db 690 GAGGCTTGTGCTGTAAGAGAGAGGCCCGACACATGGGTA 650

RESULT	5
AV072177/c	
LOCUS	
DEFINITION	
AV072177	287 bp mRNA
AV072177 Mus musculus stomach C57BL/6J adult Mus musculus CDNA clone 2200002KK2, mRNA sequence.	EST 24-JUN-1999

VERSION	AV072177.1	GI:5192005
KEYWORDS	EST.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	

REFERENCE  
AUTHORS  
Carninci, P., Shibata, K., Ozawa, Y., Kono, H., Itoh, M., Aizawa, K.,

TITLE  
 JOURNAL  
 COMMENT  
 Unpublished (1999)  
 Contact: Chile Owa  
 RIKEN Mouse ESFS  
 Aashita, S., Akiyama, T., Fukuda, S., Fukunishi, Y., Funayama, T., Hatae,  
 A., Hasegawa, N., Horii, F., Ishikawa, T., Itoh, M., Iwano, M., Kawai, D.,  
 Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuma, H., Oda, H.,  
 Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara,  
 Y., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y., Tomioka, N.,  
 Watanabe, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M.,  
 Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Genome Science Laboratory  
RIKEN  
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan  
Tel: 81-298-36-9145

Email: [genome-res@rc.riken.go.jp](mailto:genome-res@rc.riken.go.jp)  
Thermostabilization and thermocactivation of thermolabile enzymes by trehalose and its application for the syntheses of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))  
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))  
Please visit our web site (<http://genome.riken.go.jp>) for further details.

FEATURES	Location/Qualifiers
source	1. .287

Query Match	12.6%	Score 32.6;	DB 10;	Length 287;
BASE COUNT	73 a	52 g	82 t	1 others
ORIGIN				



Query Match	12.4%	Score 32;	DB 10;	Length 283;
Best Local Similarity	43.3%	Pred. No. 16;		
Matches	58;	Conservative	19;	Mismatches 55;
				Indels 2;
				Gaps 1;

RESULT	8
AL568674	
LOCUS	AL568674 1022 bp mRNA
DEFINITION	AL568674 LTI_FLU02_P1 Homo sapiens cDNA clone CS0DE004YE14 3 prime
	EST 16-FEB-2001

/note="Organ: placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the

Query Match	12.2%	Score 31.6	DB 10	Length 1022
Best Local Similarity	42.8%	Pred. No. 29		
Matches	65	Conservative 12	Mismatches 75	Indels 0
				Gaps 0

RESULT	9
LOCUS	BB500470/c
DEFINITION	BB500470 321 bp mRNA EST BB500470 RIKEN full-length enriched, 0 day neonate kidney Mus musculus cDNA clone D630028G21 3', mRNA sequence.

REFERENCE  
1 (bases 1 to 321)  
AUTHORS  
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci

MATSUZAKI, H., NOLLE, F., ISHII, T., ISHIKAWA, G., ISHIKAWA, J., ILOH, M.,  
 IZAWA, M., KADOTA, K., KAGAWA, I., KAI, C., KAWA, J., KIKUCHI, N.,  
 KIYOSAWA, H., KOJIMA, Y., KONDO, S., KOYA, S., KUNIHARA, C., KUSAKABE, M.,  
 MATSUYAMA, T., MIKI, R., MINZO, Y., NAKAMURA, M., ODA, H., OKABATA, K.,  
 ONO, T., OWA, C., SATO, H., SAKAI, C., SATO, K., SHIBATA, K., SHIBATA,  
 Y., SHIEMOTO, Y., SHINAGAWA, A., SHIRAI, T., SOGABE, Y., SUGIHARA, Y.,  
 SUZUKI, H., SUZUKI, H., TAGAWA, A., TAKAHASHI, F., TOMINAGA, N., TOYA,  
 T., TSUNODA, Y., WATANAKI, A., WATANABE, S., YAMAMURA, T., YAMANAKA, I.,  
 YANO, R., YASUNISHI, A., YOKOTA, T., YOSHIDA, K., YOSHIKI, A., YOSHINO,  
 M., MURAMATSU, M., and HAYASHIZAKI, Y.  
 RIKEN ESTS (Konno, H., et al.)  
 TITLE  
 JOURNAL  
 COMMENT  
 Unpublished (2000)  
 Contact: Yoshihide Hayashizaki

Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-resgsc.riken.go.jp/  
URL: <http://genome.gsc.riken.go.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki,  
N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermotranscription and thermotranscription of thermolabile enzymes by  
thermolabile and its application for the synthesis of full length  
cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 95 (2), 520-524 (1998)  
Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki,  
Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. *Genome Res.* 9 (5), 463-470 (1999)











